



TESIS DOCTORAL

**Análisis de los patrones de variación genética
de las poblaciones de *Giardia* que parasitan
humanos en el Área Sanitaria de Santiago de
Compostela (Galicia)**

Memoria que presenta, para optar al grado de Doctor:

Luis Basilio Gabín García

Centro de Investigación en Medicina Molecular e Enfermedades Crónicas

Facultade de Medicina e Odontoloxía

Universidade de Santiago de Compostela

Santiago de Compostela

2017



Dr. Xulio M. Maside Rodríguez, Profesor do Departamento de Ciencias Forenses, Anatomía Patolóxica, Xinecoloxía e Obstetricia e Pediatría da Universidade de Santiago de Compostela e

Dra. Carolina Bartolomé Husson, investigadora contratada da Universidade de Santiago de Compostela

INFORMAMOS

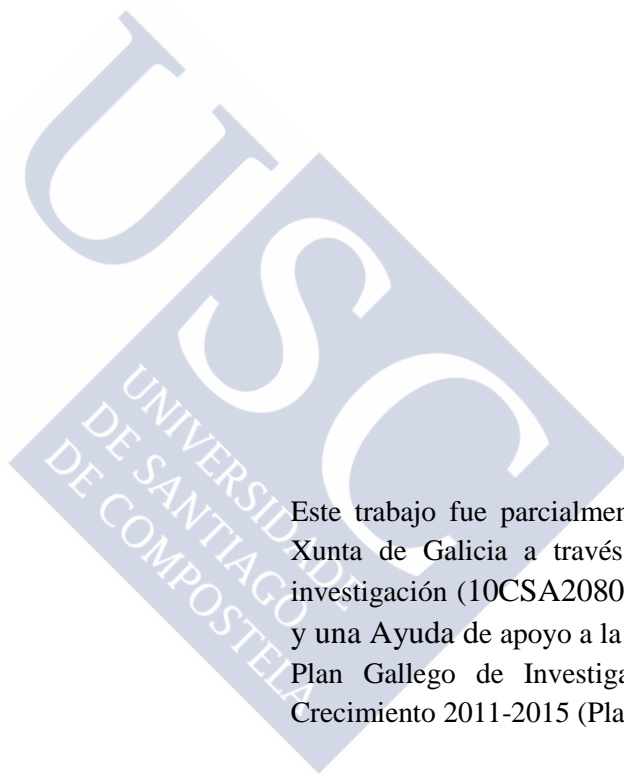
Que esta memoria titulada: “Análisis de los patrones de variación genética de las poblaciones de *Giardia* que parasitan humanos en el Área Sanitaria de Santiago de Compostela (Galicia)” e realizada por D. Luis Basilio Gabín García para optar ao Grao de Doctor, foi realizada baixo a nosa dirección, consederámola rematada e autorizamos a súa presentación perante o Tribunal calificador

E para que así conste a todos os efectos, asinamos en Santiago de Compostela, a 28 de Febreiro de 2017.

Asdo.: Xulio M. Maside Rodríguez

Asdo.: Carolina Bartolomé Husson





Este trabajo fue parcialmente financiado por la Xunta de Galicia a través de un proyecto de investigación (10CSA208038PR) a X. Maside y una Ayuda de apoyo a la etapa predoctoral del Plan Gallego de Investigación, Innovación y Crecimiento 2011-2015 (Plan I2C) a L. B. Gabín.



AGRADECIMIENTOS

En primer lugar quiero agradecer a Xulio y a Carolina por haberme acogido en su laboratorio y haber dirigido esta tesis como lo han hecho. Gracias por el trato que he recibido estos cinco años, la cercanía, el buen humor en el laboratorio todos y cada uno de los días, la libertad que he sentido para discutirlo todo, incluso cuando desde el conocimiento y la experiencia ya sabíais de salida que yo estaba más que equivocado, pero sobre todo quiero agradecer la confianza que he sentido incluso cuando no salía nada.

A José Luis, en gran parte responsable del buen ambiente del laboratorio, siempre dispuesto a echarme una mano, es muy fácil adaptarse a un laboratorio nuevo cuando dentro hay gente así.

Al Dr. José Llovo Taboada por su apoyo desde el Servizo de Microbioloxía e Parasitoloxía do Complexo Hospitalario Universitario de Santiago (SERGAS), por las muestras, por su pasión por la parasitología y la biología en general, porque siempre ha tenido un segundo para resolver una duda o enseñarme algo nuevo. Gracias también a Santiago Méndez Torres por su ayuda en mis visitas al hospital.

A la gente del servicio de secuenciación del hospital. A Bea, por su consejo en el diseño del experimento de secuenciación masiva; a Jorge por su ayuda en el tratamiento informático de los datos, y a Susana por cuidarme en su laboratorio.

Al Dr. Jose Manuel Leiro, por dejarnos utilizar su laboratorio para la micromanipulación de los quistes.

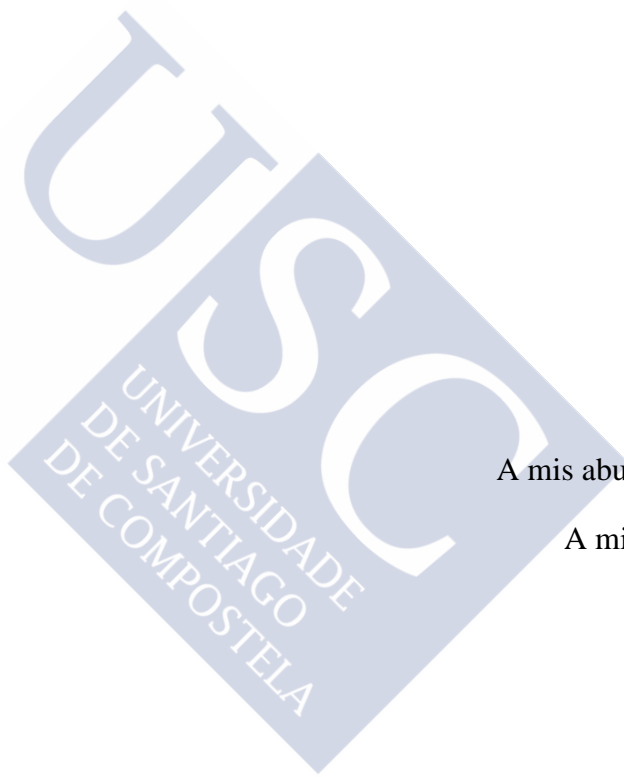
A mis padres, porque sin vosotros nada de esto sería posible, habéis sido un ejemplo a seguir durante todos estos años y mi mayor apoyo. A mi hermana, a mis abuelos, y a toda mi familia, que siempre habéis respetado y valorado los pasos que he dado.

A Seo, el mejor compañero de piso que podría haber tenido durante estos años. A María, por la carrera, por el máster, por el doctorado, por estar pendiente en este camino en paralelo.

A mi gente del colegio mayor, que me acogieron en Santiago, y se convirtieron en amigos para la vida. A los que habéis estado aquí los años de la tesis: Seo, Judo, Ada, Pati, Viri,...; y los que os habéis ido lejos y me toca echaros de menos: Lucas, Luisa, Kar, Wences,... Y a la gente que me llevo del paso por la carrera y el máster: a Laura Vidal por espabilarme, Patricia, Laura Martínez, Jaime, Cynthia, ...

A la gente del baloncesto, a mis maestros que se han convertido en amigos sobre todo a Espi y estos últimos años Pablo, por convertir el pabellón de Fontecarmoa en mi lugar de escape.





A mis abuelas, a mis abuelos,

A mi madre, a mi padre,

A mi hermana



ÍNDICE

ÍNDICE	XI
1 INTRODUCCIÓN	3
1.1 ENFERMEDADES DIARREICAS: GIARDIASIS	3
1.1.1 <i>Sintomatología y desarrollo de la enfermedad</i>	4
1.1.2 <i>Epidemiología</i>	5
1.2 CARACTERÍSTICAS MORFOLÓGICAS Y CICLO VITAL	7
1.2.1 <i>Trofozoíto</i>	7
1.2.2 <i>Quiste</i>	9
1.2.3 <i>Ciclo vital</i>	10
1.3 TAXONOMÍA Y NOMENCLATURA	13
1.3.1 <i>Filogenia</i>	13
1.3.2 <i>Taxonomía</i>	14
1.3.3 <i>Subestructuración genética: Ensamblajes</i>	15
1.3.4 <i>Variabilidad dentro de ensamblajes</i>	18
1.4 GENOMA	21
1.5 EPIDEMIOLOGÍA MOLECULAR	23
1.5.1 <i>Determinación de ensamblaje</i>	24
1.5.2 <i>Diversidad genética</i>	26
1.5.3 <i>Potencial zoonótico</i>	27
1.5.4 <i>Recombinación</i>	29
1.5.5 <i>Análisis de quistes individuales</i>	31
2 OBJETIVOS	35
3 MATERIAL Y MÉTODOS	39
3.1 EXPERIMENTO I: ANÁLISIS DE LOS PATRONES DE DIVERSIDAD GENÉTICA EN AISLADOS DE LOS ENSAMBLAJES A Y B	39
3.1.1 <i>Amplificación, clonación y secuenciación</i>	39
3.1.1.1 <i>Origen de las muestras y extracción de ADN</i>	39
3.1.1.2 <i>Selección de genes y diseño de cebadores</i>	39
3.1.1.3 <i>Determinación del ensamblaje y selección de muestras</i>	41
3.1.1.4 <i>Amplificación</i>	43
3.1.1.5 <i>Comprobación de los ensamblajes con enzimas de restricción</i>	44
3.1.1.6 <i>Clonación</i>	44
3.1.1.7 <i>Secuenciación</i>	45
3.1.2 <i>Análisis genético-poblacional</i>	45
3.1.2.1 <i>Diversidad y divergencia</i>	45
3.1.2.2 <i>Estructura poblacional</i>	46
3.1.2.3 <i>Recombinación</i>	47
3.1.2.4 <i>Análisis filogenético</i>	48
3.1.2.5 <i>Redes de haplotipos</i>	48
3.2 EXPERIMENTO II: GENOTIPADO MULTILOCUS DE QUISTES INDIVIDUALES	49
3.2.1 <i>Aislamiento de quistes individuales</i>	49
3.2.1.1 <i>Selección de muestras</i>	49
3.2.1.2 <i>Purificación y concentración de los quistes</i>	49
3.2.1.3 <i>Micromanipulación</i>	50
3.2.2 <i>Amplificación directa multiplex</i>	51
3.2.2.1 <i>Selección de marcadores y diseño de cebadores</i>	51

3.2.2.2	PCR multiplex.....	52
3.2.3	<i>Secuenciación masiva en paralelo</i>	53
3.2.3.1	Construcción de librerías de amplicones.....	54
3.2.3.2	Preparación del molde de secuenciación.....	56
3.2.3.3	Secuenciación.....	57
3.2.4	<i>Análisis de secuencias</i>	58
3.2.4.1	Identificación de alelos	58
3.2.4.1.1	Alineamiento de secuencias únicas (método USEQ)	59
3.2.4.1.2	AMPLIAN	61
3.2.4.1.3	AMPLISAS	61
3.2.4.1.4	Confirmación por Sanger.....	62
3.2.4.2	Estima de la dosis alélica	63
3.2.4.3	Análisis filogenético.....	63
4	RESULTADOS.....	67
4.1	EXPERIMENTO I: ANÁLISIS DE LOS PATRONES DE DIVERSIDAD GENÉTICA DE <i>G. DUODENALIS</i> EN AISLADOS DE PACIENTES	67
4.1.1	<i>Determinación de los ensamblajes y selección de las muestras</i>	67
4.1.2	<i>Amplificación, clonación y secuenciación de los loci gdh, bg y calt</i>	67
4.1.3	<i>Análisis de diversidad genética</i>	69
4.1.3.1	Diversidad sinónima.....	69
4.1.3.2	Diversidad no sinónima.....	72
4.1.3.3	Análisis de la varianza molecular.....	73
4.1.3.4	Diversidad haplotípica	74
4.1.3.5	Análisis filogenético.....	75
4.1.3.6	Redes de haplotipos.....	79
4.1.3.7	Recombinación.....	82
4.1.3.8	Divergencia	83
4.2	EXPERIMENTO II: GENOTIPADO MULTILOCUS DE QUISTES INDIVIDUALES DE <i>G. DUODENALIS</i> MEDIANTE SECUENCIACIÓN MASIVA EN PARALELO	84
4.2.1	<i>Purificación de quistes</i>	84
4.2.2	<i>Amplificación multiplex de quistes aislados</i>	85
4.2.2.1	Optimización de la PCR multiplex.....	85
4.2.2.2	Control de la amplificación de ADN libre	85
4.2.2.3	PCR directa multiplex de quistes individuales.....	88
4.2.3	<i>Resultados de la secuenciación masiva en paralelo</i>	89
4.2.3.1	Rendimiento de la reacción de secuenciación	89
4.2.4	<i>Genotipado multilocus</i>	91
4.2.4.1	Identificación de alelos	92
4.2.4.1.1	Identificación de artefactos	93
4.2.4.1.2	Confirmación por secuenciación Sanger.....	99
4.2.4.2	Genotipos	100
4.2.4.3	Heterozigosidad	103
4.2.4.4	Dosis alélica	104
4.2.4.5	Genotipos multilocus	106
4.2.4.6	Análisis filogenético.....	107
5	DISCUSIÓN	115
5.1	DIVERSIDAD GENÉTICA EN LOS ENSAMBLAJES A Y B.....	118
5.2	DIVERSIDAD GENÉTICA DENTRO DE LOS AISLADOS	122
5.3	AMPLIFICACIÓN Y SECUENCIACIÓN DE QUISTES INDIVIDUALES	124

5.4 DETERMINACIÓN DE LOS GENOTIPOS.....	126
5.5 HETEROZIGOSIDAD <i>vs.</i> INFECCIONES MIXTAS	127
5.6 INTERCAMBIO GENÉTICO	129
5.7 FILOGENIA	133
6 CONCLUSIONES	137
7 BIBLIOGRAFÍA	141
8 ANEXOS.....	159
8.1 TABLAS DE HAPLOTIPOS DEL EXPERIMENTO I.....	159
8.2 TABLAS DE HAPLOTIPOS DEL EXPERIMENTO II	164
8.3 NÚMERO DE LECTURAS POR ALELO.....	167
8.4 FIGURAS DE LAS FRECUENCIAS ALÉLICAS RELATIVAS EN LOS QUISTES INDIVIDUALES DE LOS AISLADOS D, E, H, K, M.....	172
8.5 ÁRBOLES FILOGENÉTICOS DEL EXPERIMENTO II	176





Abreviaturas

A.N. – número de acceso (*Accession Number*)
ADN – Ácido desoxirribonucleico
AMOVA – Análisis de varianza molecular (*Analysis of Molecular Variance*)
ARN – Ácido ribonucleico
ASH – Heterozigosidad alélica (*Allelic sequence heterozygosity*)
bg – beta-giardin
BLAST – *Basic Local Alignment Search Tool*
calt – Caltractin
CDC – *Centre for Disease Control and Prevention*
CWPs – proteínas de la pared del quiste (*Cyst Wall Proteins*)
DMSO – Dimetilsulfóxido
dNTP – desoxinucleótido trifosfato
e.g. – *exempli gratia* (por ejemplo)
ECDC – *European Center for Disease Prevention and Control*
EDO – enfermedad de declaración obligatoria
EE – Error estándar
EEUU – Estados Unidos
ef-1 – *elongation factor 1*
emPCR – PCR en emulsión
ENA – European Nucleotide Archive
ens. – ensamblaje
ESVs – Vesículas específicas de la enquistación (*encystation specific vesicles*)
et al. – et alii (y colaboradores)
fd – *ferredoxin*
gdh – *glutamate dehydrogenase*
hA – haplotipo del ensamblaje A
hB – haplotipo del ensamblaje B
IGV – Integrative Genome Viewer
indel – inserción o delección
ISP – *Ion Sphere™ Particle*

JC – Jukes-Cantor
kB – Kilobases
Mb – Megabases
MLG – Genotipo multilocus (*Multilocus genotype*)
mlh1 – proteína *mlh1*
MPS – secuenciación masiva en paralelo (*Massive Parallel Sequencing*)
NCBI – *National Center for Biotechnology Information*
NJ – *Neighbor-Joining*
OMS – Organización Mundial de la Salud
pb – pares de bases
PBS – Tampón fosfato salino (*Phosphate Buffered saline*)
PCR – Reacción en cadena de la polimerasa (*Polymerase Chain Reaction*)
PGM – *Personal Genome Machine*
RAPD – *Random Amplification of Polymorphic DNA*
RENAVE – Red Nacional de Vigilancia Epidemiológica
RFLP – *Restriction Fragment Length Polymorphism*
rpL7a – *ribosomal protein L7a*
Shorah – *Short Reads Assembly into Haplotypes*
SIM – Sistema de Información Microbiológica
SNV – Variantes de un único nucleótido (*Single Nucleotide Variant*)
SSUrRNA – *Small Subunit of ribosomal RNA*
TBE – Tris-Bórico-EDTA
TN – Tamura-Nei
tpi – *triose phosphate isomerase*
UE – Unión Europea
USEQ – secuencias únicas (*Unique sequences*)
UTR – Regiones no traducidas (*untranslated regions*)
vs. – *versus*
VSP – proteínas variantes de superficie (*variant surface proteins*)
WHO – *World Health Organization*
ZOOPNET – *Zoonotic protozoan Network*



1. Introducción



1 Introducción

El género *Giardia* engloba un grupo de protozoos flagelados que parasitan distintas especies de vertebrados causando problemas intestinales, principalmente diarrea. Su presencia en agua para consumo u ocio puede provocar la aparición de brotes epidémicos en humanos, así como pérdidas económicas en el sector ganadero. Se trata de parásitos presentes prácticamente en todo el planeta, muy comunes tanto en áreas castigadas por la pobreza como en países industrializados (Adam, 2001; Ankarklev *et al.*, 2010; Thompson y Monis, 2011).

El microscopista Antony van Leeuwenhoek fue el primero en notificar la observación de estos parásitos en 1681 mientras investigaba sus propias heces en busca del causante de su diarrea. Los describió como “animáculos” alargados dotados de “pies” que les permitían moverse lentamente. A mediados del siglo XIX, Vilem Dusan Lambl ofreció la primera descripción detallada de este flagelado y propuso el nombre de *Cercomonas intestinalis*. Este nombre fue enseguida rechazado por la comunidad científica, ya que además de incurrir en homonimia con otra especie previamente descrita, la asignación del género era incorrecta. En 1882, J. Kunstler utilizó por primera vez el nombre genérico *Giardia* para describir un protozoo que había encontrado en el intestino de renacuajos (*Giardia agilis*) y que pertenecía al mismo género que el parásito encontrado en humanos (Thompson y Monis, 2011). En 1987 se demostraron los postulados de Koch para la especie que parasita humanos: *Giardia duodenalis* (Nash *et al.*, 1987).

1.1 ENFERMEDADES DIARREICAS: GIARDIASIS

La diarrea es una de las patologías más comunes en todo el mundo, con especial importancia en la población infantil menor de cinco años en la que se sitúa como segunda causa de muerte (760 000 muertes anuales a nivel mundial; World Health Organization, 2013). Se presenta en la mayor parte de los casos como síntoma de una infección intestinal que puede estar causada por distintos tipos de virus (rotavirus, enterovirus, etc.), bacterias (*Escherichia coli*, *Shigella sp.*, *Campylobacter sp.*, etc.) o parásitos eucariotas (*G. duodenalis*, *Cryptosporidium sp.*, *Entamoeba histolytica*, etc.). *G. duodenalis* está reconocido como uno

de los mayores causantes de diarreas de origen parasitario tanto en países industrializados como en países no industrializados (Fletcher *et al.*, 2012).

1.1.1 Sintomatología y desarrollo de la enfermedad

La infección por *Giardia* se produce exclusivamente tras la ingesta de quistes, ya sea por beber o tragar accidentalmente agua contaminada, por consumir alimentos contaminados o mediante contacto directo entre personas en condiciones de higiene insuficientes. La contaminación de sistemas de agua potable puede provocar brotes epidémicos (Castro-Hermida *et al.*, 2008; revisado en Efstratiou *et al.*, 2017). El mayor brote reciente ocurrió en Noruega en 2004 con más de 1300 casos confirmados (Nygård *et al.*, 2006). Además, *G. duodenalis* presenta potencial zoonótico, ya que puede infectar tanto a humanos como a otras especies de mamíferos y aves –algunas de ellas domesticadas, como mascotas o ganado– que podrían servir como reservorio del parásito. Sin embargo, el papel que juega la transmisión entre humanos y animales no está totalmente claro (ver apartado 1.5.3) (revisado en Ryan y Cacciò, 2013).

La capacidad del parásito para infectar al hospedador y colonizar el intestino delgado depende de factores propios de la cepa infectante y del hospedador. Se trata de infecciones típicamente autolimitantes en el caso de hospedadores inmunocompetentes, pero que pueden presentar manifestaciones clínicas muy diversas que van desde asintomáticas a sintomáticas agudas o crónicas, que en los casos más severos no responden al tratamiento (Ankarklev *et al.*, 2010; Robertson *et al.*, 2010; Thompson y Monis, 2011). Los factores que influyen en la manifestación clínica de la enfermedad no están claros, pudiendo estar implicados factores de virulencia propios del parásito, la interacción con otros enteropatógenos y factores propios del hospedador como la edad, el perfil genético, el sistema inmune, la microbiota normal o la dieta (Ankarklev *et al.*, 2010; Thompson y Monis, 2011; Einarsson *et al.*, 2016).

Los síntomas aparecen a la semana o dos semanas de la ingesta de los quistes y los más destacados son diarrea, flatulencia, náuseas, calambres intestinales, dolor abdominal, pérdida de peso y malabsorción (Adam, 2001; Ankarklev *et al.*, 2010; Thompson y Monis, 2012). No existe consenso acerca del mecanismo por el cual se producen estos síntomas y se desarrolla la enfermedad (patogénesis), ya que las células de *Giardia* no son invasivas ni secretan

ninguna toxina conocida. Actualmente, la giardiasis se considera como una enfermedad multifactorial en la que participan numerosos mecanismos inducidos por el parásito, entre los que estarían el aumento de la permeabilidad intestinal (causado por la inducción de la apoptosis y la parada del ciclo celular de las células epiteliales, así como por la disrupción de las uniones estrechas entre células epiteliales), el aumento de la secreción de Cl^- , el acortamiento de las microvellosidades, la alteración de la proporción vellosidad/cripta, la hipermotilidad intestinal, la hiperplasia de la cripta, el aumento de secreción de mucosa y la inducción de cambios en la microbiota normal (Ankarklev *et al.*, 2010; Thompson y Monis, 2011; Einarsson *et al.*, 2016).

Además de los síntomas propios de la infección, en los últimos años se ha relacionado la giardiasis con la aparición de secuelas crónicas después de que el parásito haya dejado de ser detectable en el hospedador. Diversos estudios han asociado la infección por *Giardia* con el desarrollo de síndrome de colon irritable, fatiga crónica, malnutrición, deterioro cognitivo y otras manifestaciones extraintestinales (tales como alergia a los alimentos, urticaria, artritis reactiva y manifestaciones oculares inflamatorias) (Naess *et al.*, 2012; Hanevik *et al.*, 2014; Bartelt y Sartor, 2015).

1.1.2 Epidemiología

Existen varias razones que dificultan la obtención de estimas fiables de la prevalencia de la giardiasis: i) su declaración no es obligatoria en muchos países, ii) los métodos diagnósticos utilizados en los distintos laboratorios no están estandarizados, iii) presenta una alta tasa de infecciones asintomáticas y iv) muchas de las personas infectadas no tienen acceso a sanidad. En este contexto, la tasa de prevalencia de la giardiasis se estima entre el 2% y el 7% en Europa, EE.UU., Canadá y Australia, pudiendo llegar hasta el 40% en las regiones con peores condiciones higiénico-sanitarias de los países no industrializados (Thompson y Monis, 2012).

En España, la giardiasis es una enfermedad de declaración obligatoria (EDO) desde el año 2015 (Orden SSI/445/2015). Hasta entonces, la declaración de esta enfermedad a través del Sistema de Información Microbiológica (SIM) era voluntaria, lo que explica la ausencia de datos para algunas comunidades autónomas en los informes de la Red Nacional de

Vigilancia Epidemiológica (RENAVE). El número de casos anuales declarados a través del SIM entre los años 2009 y 2013 osciló entre 350 y 900, aunque tan solo 26 laboratorios notificaron de forma constante durante este periodo. La inclusión en 2014 de la giardiasis como EDO en algunas comunidades autónomas disparó hasta 1483 los casos totales declarados ese año a pesar de que no se dispuso de datos de comunidades como Cataluña, Madrid o Galicia (Centro Nacional de Epidemiología, 2016). En 2015, se notificaron desde España 1627 casos a la Agencia Europea para el Control de las Enfermedades Infecciosas (*European Centre for Disease Prevention and Control*, ECDC), lo que supuso un incremento del 100% desde 2012 (ECDC, 2017). Es predecible que este número aumente en los próximos informes con la aplicación de la nueva legislación a nivel estatal (mencionada al inicio del párrafo), permitiendo obtener estimas más fiables acerca de la prevalencia de la giardiasis.

La ECDC registró entre 16 000 y 18 000 casos anuales desde 2007 (ECDC, 2017). Sin embargo, los protocolos de notificación dependen de cada país miembro y pueden cambiar incluso de año en año, produciendo grandes variaciones en el número de casos. La última encuesta epidemiológica (2015) solo incluyó datos de giardiasis de 24 países, dejando fuera a algunos de los de mayor población como Francia, Italia, Países Bajos, o Grecia (ECDC, 2017). Por tanto, estas cifras no son representativas de la importancia de la giardiasis en Europa a nivel cuantitativo, aunque sí tienen valor cualitativo al permitir establecer tendencias anuales o patrones epidemiológicos que podrían ser extrapolables al conjunto del continente.

En el año 2015 se notificaron 18 012 casos en el Área Económica Europea, lo que supuso un incremento respecto a años anteriores debido principalmente a un aumento del 25% en el número de casos registrados en Reino Unido (4536 vs. 3628 en 2014; ECDC, 2017). De éstos necesitaron hospitalización un total de 401, aunque la cifra podría ser mayor, ya que algunos países no proporcionaron este dato. Los casos notificados por cada 100 000 habitantes variaron entre 0,25 en Portugal (primer año del que hay datos para este país) y 17,29 en Bulgaria (ECDC, 2017).

En EE.UU., el Centro de Control de Enfermedades (CDC), recibió la notificación de una media de 20 000 casos al año (incidencia: ~7,5/100 000 habitantes) durante la década de los 2000, sufriendo un descenso en los siguientes años hasta aproximadamente 15 000 casos anuales (~6/100 000 habitantes) (Hlavsa *et al.*, 2005; Yoder y Beach, 2007; Yoder *et al.*, 2010; Yoder *et al.*, 2012; Painter *et al.*, 2015). El coste estimado que suponen las

hospitalizaciones por giardiasis en EEUU es de 34 millones de dólares anuales (Collier *et al.*, 2012). En Asia, África y América latina, dónde existe menos información al respecto, la Organización Mundial de la Salud (OMS) daba en su informe del año 1996 una estima de 200 millones de personas con giardiasis sintomática con 500 000 nuevos casos registrados cada año (World Health Organization, 1996).

Diversos estudios sugieren que los datos aportados por las agencias de vigilancia epidemiológica suponen en su mayor parte una infraestima de la incidencia real de la giardiasis en la población. En un meta-análisis de estudios epidemiológicos en los países nórdicos, se estimó que por cada caso registrado ocurrían 867 casos sintomáticos sin registrar en Finlandia, 634 en Noruega y 250 en Suecia (uno de los países de la UE con más casos declarados) (Hörman *et al.*, 2004).

Tanto en Europa como en EEUU, se aprecia cierta estacionalidad con una mayor incidencia entre el final del verano y el otoño, aunque el número de casos registrados es bastante constante durante todo el año. El incremento de casos en los meses estivales en el hemisferio norte podría estar relacionado con un aumento de las actividades sociales o recreativas al aire libre, como el baño en piscinas públicas, ríos o lagos (Lal *et al.*, 2012).

1.2 CARACTERÍSTICAS MORFOLÓGICAS Y CICLO VITAL

Durante su ciclo vital, *G. duodenalis* se diferencia en dos formas celulares distintas: el trofozoíto, que es la forma reproductiva y provoca la enfermedad al colonizar el intestino, y el quiste, que es la forma infectiva, capaz de sobrevivir en el medio externo al hospedador y de causar una nueva infección tras la ingesta (Adam, 2001; Ankarklev *et al.*, 2010).

1.2.1 Trofozoíto

El trofozoíto es una célula piriforme (forma de pera partida por la mitad, en que la parte plana es ventral) de entre 12-15 μm de largo y 5-9 μm de ancho. Presenta cuatro pares de flagelos (anterior, posterior, caudal y ventral) y una estructura ventral característica llamada disco adhesivo, que permite su adhesión a las células epiteliales del intestino (Ankarklev *et al.*, 2010). Entre los elementos que forman parte del citoesqueleto, además de los flagelos y el

disco ventral, destaca el cuerpo medio, que está compuesto fundamentalmente por microtúbulos situados de forma transversal al eje antero-posterior y es claramente observable al microscopio. Su variabilidad morfológica le ha valido un papel clave en la identificación de las distintas especies o grupos morfológicos dentro del género, a pesar de que su función es desconocida (Figura 1.1a) (Thompson y Monis, 2011).

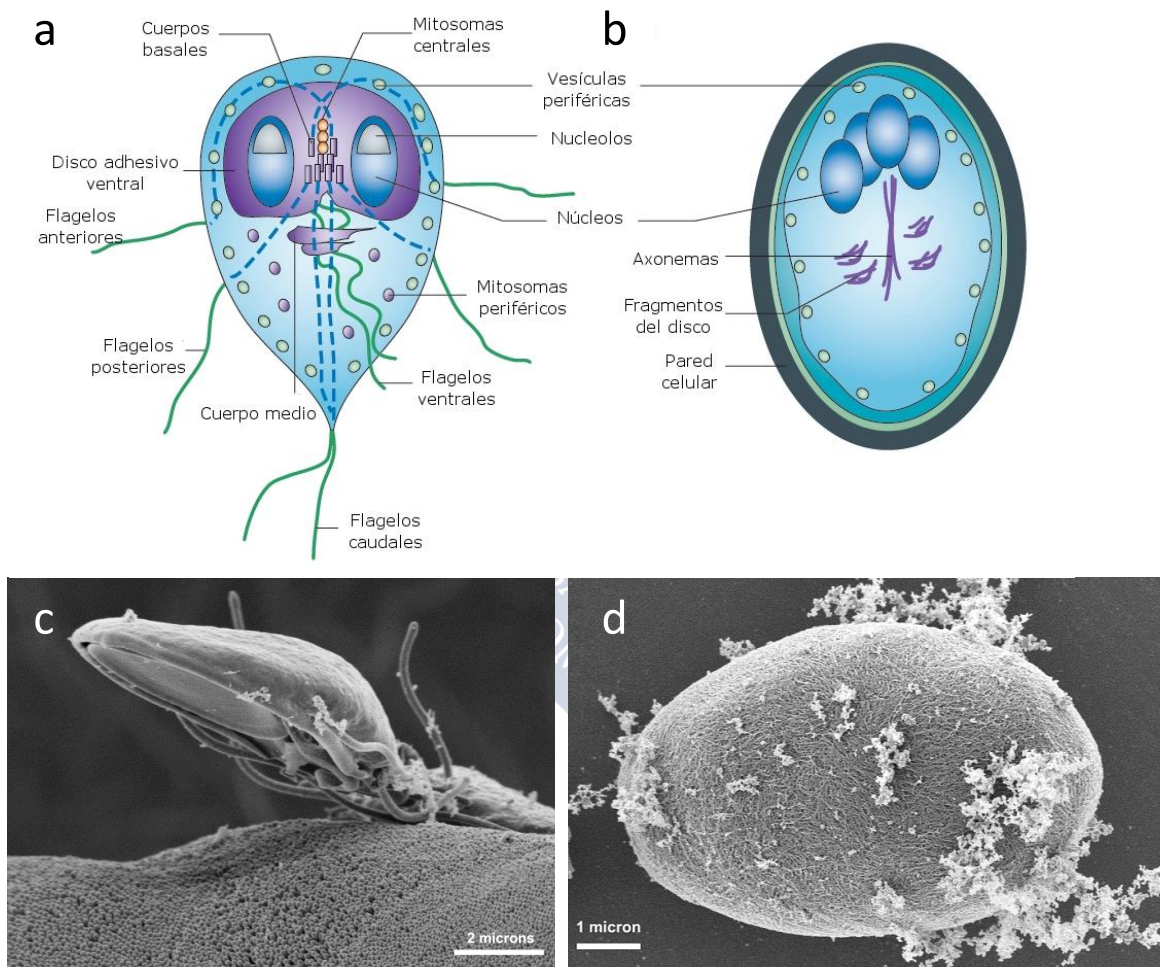


Figura 1.1: Características morfológicas de las formas celulares de *G. duodenalis*. a) representación esquemática del trofozoíto y sus orgánulos, b) representación esquemática del quiste y sus orgánulos. Modificado de Ankarklev *et al.* (2010); c) imagen de microscopía electrónica de barrido de un trofozoíto a punto de adherirse al epitelio intestinal con su disco ventral; d) imagen de microscopía electrónica de barrido de un quiste. Estas imágenes se encuentran disponibles para uso público en la *Public Health Image Library* (PHIL), <https://phil.cdc.gov/phil/details.asp>, ID#:11630 y ID#: 11642, respectivamente.

La célula de *Giardia* muestra importantes diferencias respecto a la célula eucariota típica: presenta dos núcleos y carece de mitocondrias, peroxisomas y aparato de Golgi propiamente dicho (Thompson y Monis, 2011). A diferencia de otras especies binucleadas, los núcleos son equivalentes: ambos son transcripcionalmente activos y cada uno de ellos contiene dos copias completas del genoma, convirtiendo a esta célula en tetraploide ($2 \times 2n$) (Kabnick y Peattie, 1990; Bernander *et al.*, 2001; Yu *et al.*, 2002).

Aunque el trofozoíto no presenta mitocondrias sí se han encontrado otros orgánulos rodeados de una doble membrana que participan en la maduración de las proteínas hierro-azufre. Estas estructuras reciben el nombre de mitosomas y están considerados como remanentes funcionales de las mitocondrias que habrían perdido su genoma y la capacidad de generar energía a través de la cadena de transporte de electrones (Tovar *et al.*, 2003). Algo similar ocurre con el aparato de Golgi. No existe ninguna estructura en el trofozoíto que se pueda identificar morfológicamente como un Golgi típico y se ha observado que el retículo endoplasmático y la envoltura nuclear llevan a cabo algunas de las funciones que realiza el aparato de Golgi en eucariotas superiores, como el empaquetamiento y la clasificación de las proteínas (Elias *et al.*, 2008). Sin embargo, durante la enquistación se forman las vesículas específicas de la enquistación (*encystation specific vesicles*, ESVs) que trasladan las proteínas de la pared (*cyst wall proteins*, CWP) desde el retículo endoplasmático hacia la superficie celular en un modelo de transporte similar al del aparato de Golgi (Marti *et al.*, 2003).

1.2.2 Quiste

El quiste es una célula ovalada de entre 8-12 μm de largo por 7-10 μm de ancho, rodeada por una pared de 0,3- 0,5 μm de grosor (Ankarklev *et al.*, 2010). La pared está formada por una doble membrana interna rodeada por una capa filamentosa compuesta por el homopolímero beta (1-3)-*N*-acetyl-*D*-galactosamina en asociación con las proteínas de pared CWP1, CWP2 y CWP3 (Gerwig *et al.*, 2002; Ankarklev *et al.*, 2010). Es altamente insoluble y proporciona al quiste una elevada resistencia frente a distintos agentes químicos. El quiste presenta cuatro núcleos tetraploides ($4 \times 4n$) que se forman por duplicación sin citocinesis de los dos núcleos del trofozoíto original seguida por una nueva ronda de replicación del genoma (Figura 1.1b) (Bernander *et al.*, 2001). Los quistes son infecciosos inmediatamente después de

su excreción con las heces, son muy estables y pueden sobrevivir semanas e incluso meses en el medio ambiente (Thompson y Monis, 2011).

1.2.3 Ciclo vital

Giardia presenta un ciclo vital monoxeno –se completa en un único hospedador–, que consta de tres etapas: i) diferenciación de quiste a trofozoíto o esquistación, ii) división mitótica de los trofozoítos y iii) diferenciación de trofozoíto a quiste o enquistación.

La infección comienza con la ingesta de un número de quistes que puede llegar a ser tan bajo como 10-25 individuos. El paso de los quistes por el ambiente ácido del estómago desencadena el proceso de diferenciación celular hacia el trofozoíto. Cuando el quiste alcanza el intestino delgado, se genera una apertura en la pared en uno de sus polos por la que sale una célula flagelada con cuatro núcleos tetraploides conocida como esquizoíto. Esta forma transitoria sufre inmediatamente dos citocinesis consecutivas sin necesidad de entrar en fase de síntesis, dando como resultado cuatro trofozoítos tetraploides ($2 \times 2n$) en un periodo muy corto de tiempo (Figura 1.2). Este podría ser uno de los factores determinantes para que *Giardia* presente una dosis infectiva mínima tan baja. Durante el proceso de división del esquizoíto aumenta el metabolismo, se segregan los orgánulos y se ensambla de nuevo el disco ventral (datos extraídos de la revisiones: Ankarklev *et al.*, 2010; Thompson y Monis, 2011).

El modo en que segregan los núcleos en estas dos divisiones determina si se mantiene el genotipo original del trofozoíto que se enquistó o se generan nuevos genotipos. En dos estudios independientes publicados en 2012 se observó que durante todo el proceso de esquistación los núcleos hermanos se mantienen asociados en pares, de modo que los cuatro trofozoítos resultantes contienen una copia de cada uno de los dos núcleos del trofozoíto original (segregación ecuacional; Figura 1.2), en contraposición a lo que ocurriría si cada trofozoíto heredase dos núcleos idénticos fruto de la duplicación de uno de los núcleos originales (segregación reduccional) (Jiráková *et al.*, 2012; Carpenter *et al.*, 2012).

Cuando los trofozoítos alcanzan el intestino delgado utilizan su disco ventral a modo de ventosa para fijarse al epitelio intestinal mediante una unión no específica. En este momento

se produce la fase proliferativa del ciclo, en la que los trofozoítos se reproducen por división mitótica semiabierta con reparto ecuacional de los núcleos en la citocinesis. En la mitosis semiabierta las envolturas nucleares no llegan a desintegrarse, sino que los microtúbulos del huso mitótico las atraviesan por unas aperturas que se forman en los polos. De nuevo, la segregación ecuacional implica que los trofozoítos resultantes reciben una copia de cada uno de los núcleos, reproduciendo el genotipo materno, sin pérdida de heterozigosidad (Figura 1.2) (Sagolla *et al.*, 2006).

En la fase G2 del ciclo celular del trofozoíto ($2 \times 4n$) existe un punto de control para la enquistación que determina si el trofozoíto entra en fase de división celular o de diferenciación en quiste en respuesta a cambios en el ambiente intestinal. Cuando los enterocitos maduros se desprenden de la pared intestinal durante el proceso de renovación constante del epitelio, los trofozoítos que se encuentran adheridos a los mismos deben desplazarse para unirse de nuevo a los enterocitos jóvenes usando su disco ventral. Si por el contrario se mantienen en el lumen, expuestos a elevados niveles de bilis, bajos niveles de colesterol y pH básico, inician el proceso de enquistación (Lauwaet *et al.*, 2007).

Durante la enquistación se producen cambios estructurales como la internalización de los flagelos y la fragmentación del disco ventral, de modo que el trofozoíto pierde la capacidad de adhesión. El metabolismo se reduce progresivamente hasta alcanzar un estado durmiente y la célula va adquiriendo una forma redondeada. Se forman las ESVs que contienen las proteínas CWP y que migran hacia la periferia de la célula reproduciendo un modelo de transporte de vesículas similar al del aparato de Golgi (Lauwaet *et al.*, 2007; revisado en Ankarklev *et al.*, 2010; Thompson y Monis, 2011).

Durante la enquistación se produce una división de los núcleos sin citocinesis (Figura 1.2). De este modo, el quiste aún inmaduro presenta dos pares de núcleos hermanos que se sitúan en polos opuestos de la célula ($4 \times 2n$). Inmediatamente después de esta división se deposita la pared del quiste y uno de los pares de núcleos migra hacia el polo opuesto. Finalmente se produce una última ronda de replicación del ADN de modo que en los quistes maduros cada núcleo contiene cuatro copias del genoma ($4 \times 4n$) (Poxleitner *et al.*, 2008; Carpenter *et al.*, 2012; Jiráková *et al.*, 2012).

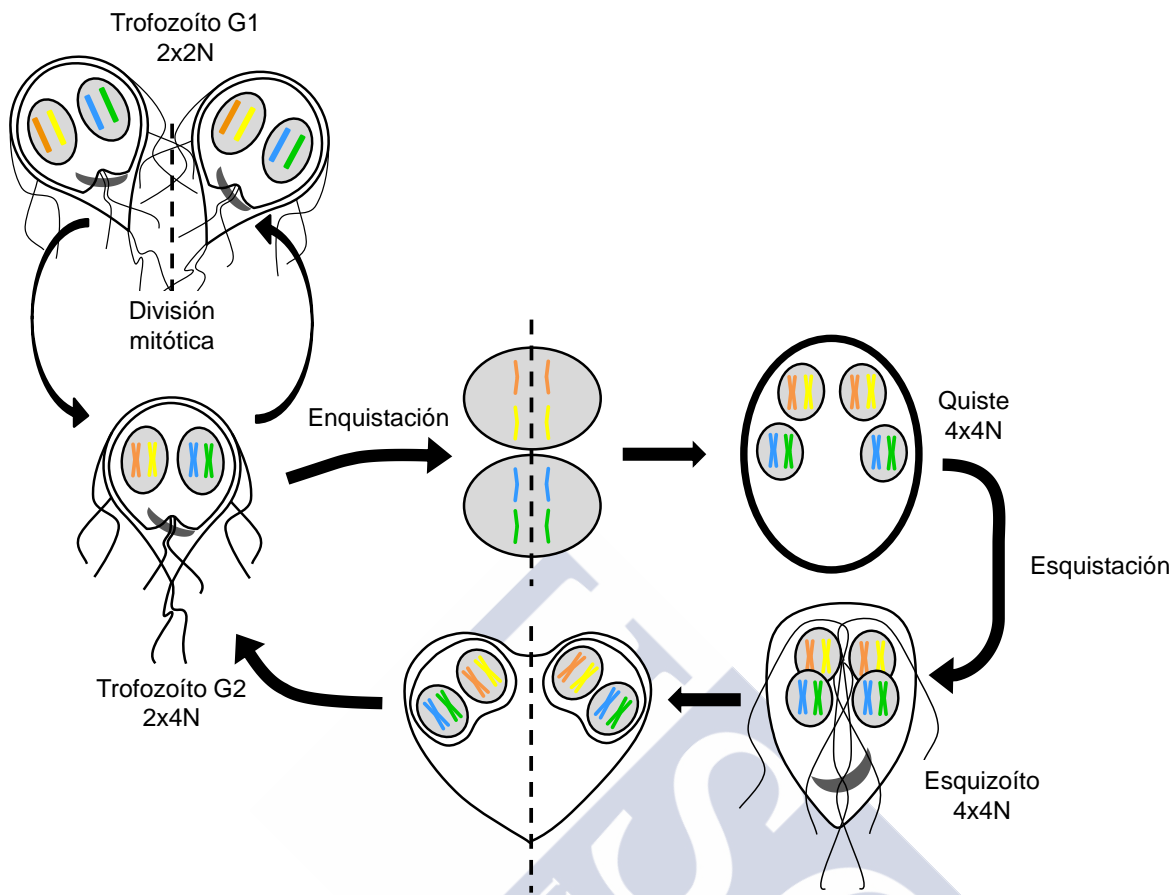


Figura 1.2: Ciclo vital de *G. duodenalis*. Basado en Jiráková *et al.* (2012). Los colores identifican los cuatro complementos genómicos. Las líneas de puntos indican segregaciones ecuacionales (mitosis, enquistación y esquistación) en las que los núcleos resultantes de la división se encuentran asociados en parejas idénticas a la de la célula original (trofozoíto G2).

En varios estudios se describió la fusión de las envolturas de los núcleos hermanos durante la enquistación (Poxleitner *et al.*, 2008; Jiráková *et al.*, 2012). Además, se detectó la transferencia de plásmidos tanto episomales (Poxleitner *et al.*, 2008) como integrados en el genoma (Carpenter *et al.*, 2012), demostrando la existencia de intercambio genético entre núcleos a través de estas comunicaciones transmembrana. Esto justificó la propuesta de un proceso de reproducción parasexual en *Giardia* mediante cariogamia e intercambio genético entre núcleos que recibe el nombre de diplomixis (Poxleitner *et al.*, 2008).

1.3 TAXONOMÍA Y NOMENCLATURA

Las numerosas peculiaridades que presentan estos protozoos y al amplio rango de hospedadores en los que fueron detectados, provocaron que tanto la posición filogenética del género *Giardia* dentro de los eucariotas, como su taxonomía hayan sido objeto de debate desde su identificación.

1.3.1 Filogenia

Debido a la ausencia en el trofozoíto de varios orgánulos complejos característicos de la célula típica eucariota (ver apartado 1.2.1), *Giardia* se clasificó inicialmente en una rama basal del árbol evolutivo de los eucariotas que se habría separado del tronco principal antes de la endosimbiosis mitocondrial (Cavalier-Smith, 1987). Los primeros análisis filogenéticos del ARN ribosómico apoyaron esta hipótesis, ya que identificaron una secuencia inusual que situó a este parásito en un linaje de divergencia temprana (Edlind y Chakraborty, 1987; Sogin *et al.*, 1989; van Keulen *et al.*, 1993). Esta posición fue apoyada por los datos de otros loci conservados como los factores de elongación y las ciclinas dependientes de kinasas (Hashimoto *et al.*, 1994; Hashimoto *et al.*, 1995; Riley y Krieger, 1995). Por estas razones *Giardia* fue considerada durante años un organismo clave para comprender el origen y la evolución de los eucariotas.

Sin embargo, este punto de vista fue perdiendo apoyo a medida que las propiedades biológicas de este protozoo se fueron conociendo más a fondo y se estudió con mayor detalle su posición filogenética respecto al resto de eucariotas (Embley y Hirt, 1998; Philippe *et al.*, 2000; Lloyd y Harris, 2002). El análisis exhaustivo de los caracteres ultraestructurales señaló a *Giardia* como uno de los miembros de su orden que más cambios experimentó en su adaptación a su vida parasitaria (Siddall *et al.*, 1992). Por ejemplo, se demostró que la ausencia de mitocondrias se debió a una pérdida secundaria como consecuencia de la adaptación al modo de vida parasitario anaerobio y no a que *Giardia* se hubiese separado del tronco principal de los eucariotas antes de la adquisición de las mismas (Hashimoto *et al.*, 1998; Horner y Embley, 2001; Tovar *et al.*, 2003).

Según la filogenia clásica, *Giardia* pertenece al Filo Sarcomastigophora, Subfilo Mastigophora (=Flagellata), Clase Zoomastigophorea, Orden Diplomonadida y Familia Hexamitidae (Vickerman, 1982; revisado en Plutzer *et al.*, 2010); siendo comúnmente referida como un Diplomonadido ya que, al igual que el resto de miembros de este orden (*Hexamita*, *Spironucleus*, *Trepomonas* y *Octomitus*), presenta la particularidad de tener dos núcleos equivalentes (Kabnick y Peattie, 1990). Sin embargo, una nueva sistemática basada en la combinación de datos genéticos y características ultraestructurales y bioquímicas, la situó en el Filo Metamonada, Subfilo Trichozoa, Superclase Eopharyngia, Clase Trepomonadea, Subclase Diplozoa, Orden Giardiida y Familia Giardiidae, separando *Giardia* y *Hexamita* en órdenes distintos (Cavalier-Smith, 2003; revisado en Plutzer *et al.*, 2010; Thompson y Monis, 2012). En este caso *Giardia* formaría parte de uno de los muchos linajes evolutivos en los que divergieron los eucariotas en sus orígenes.

1.3.2 Taxonomía

La controversia alrededor de la nomenclatura dentro del género es mayor aún debido al amplio espectro de hospedadores y a la ausencia de diferencias morfológicas entre las formas observadas en cada hospedador. Actualmente se utilizan indistintamente los nombres *G. duodenalis*, *G. lamblia* y *G. intestinalis* para referirse a las formas de *Giardia* que parasitan humanos. Sin embargo, para algunos autores solo el primero sería correcto de acuerdo con las reglas de la nomenclatura zoológica (Thompson y Monis, 2011). El nombre genérico debe ser el utilizado para describir la forma del parásito en anfibios: *Giardia agilis*, Kunstler 1882; ya que el género asignado en las descripciones anteriores de este protozoo se reveló incorrecto: *Cercomonas intestinalis*, Lambl 1859 en humanos y *Hexamita duodenalis*, Davaine 1875 en conejos. Para el nombre específico de la forma que parasita humanos prevalece *duodenalis* sobre *intestinalis*, ya que *Cercomonas intestinalis* incurre en homonimia con otra especie y nunca debió ser propuesto. Esto es así siempre que la especie observada por Davaine en conejos sea la misma que la que parasita humanos. Si se llegase a demostrar lo contrario sería necesario renombrar la especie o especies que parasitan humanos (Thompson y Monis, 2011).

Durante un tiempo existió la tendencia a asignar nuevas especies en función del hospedador, y se llegaron a describir hasta 51 especies dentro del género. Basándose en criterios morfológicos, Filice (1952) las agrupó en tres grupos a los que otorgó el rango de

especie: *G. duodenalis*, *G. muris* y *G. agilis*. A esta clasificación, ampliamente aceptada en la taxonomía actual, se han incorporado tres nuevas especies que presentan diferencias morfológicas con respecto a las anteriores: *G. psittaci*, *G. ardeae* y *G. microti* (Tabla 1.1) (Thompson y Monis, 2011).

Tabla 1.1: Características morfológicas y hospedadores de las especies del género *Giardia*.

Especie	Hospedador	Características morfológicas	Tamaño del trofozoíto (μm)
<i>G. duodenalis</i>	Amplio rango de mamíferos, incluyendo humanos	Trofozoítos piriformes con cuerpos medios en forma de garra	12-15 / 6-8
<i>G. agilis</i>	Anfibios	Trofozoítos largos y estrechos con cuerpos medios en forma de maza/porra	20-30 / 4-5
<i>G. muris</i>	Roedores	Trofozoítos redondeados con cuerpos medios pequeños y redondeados	9-12 / 5-7
<i>G. ardeae</i>	Aves	Trofozoítos redondeados con una muesca prominente en el disco ventral y flagelo caudal rudimentario. Cuerpos medios de redondeados-ovales a forma de garra	~10 / ~6.5
<i>G. psittaci</i>	Aves	Trofozoítos piriformes con cuerpos medios en forma de garra, pero sin borde ventro-lateral	~14 / ~6
<i>G. microti</i>	Roedores	Trofozoítos piriformes con cuerpos medios en forma de garra. Los quistes maduros contienen trofozoítos totalmente diferenciados	12-15 / 6-8

Nota: Adaptada de la Tabla 1.1 de Thompson y Monis (2011)

1.3.3 Subestructuración genética: Ensamblajes

Durante años la ausencia de caracteres morfológicos variables impidió establecer divisiones dentro de los grupos morfológicos de Filice, aunque el propio autor reconoció que la posición taxonómica de estas categorías morfológicas debería ser revisada cuando nuevas técnicas con mayor resolución lo permitiesen (Filice, 1952). Los avances de la biología han posibilitado el estudio de *G. duodenalis* a nivel molecular, descubriendo la existencia de grandes diferencias genéticas entre aislados (Nash y Keister, 1985; e.g., Andrews *et al.*, 1989; Mayrhofer *et al.*, 1995; Monis *et al.*, 1998; Monis *et al.*, 1999).

En 1985, se describieron los grupos 1, 2 y 3 en base a los patrones de restricción del genoma, los patrones de hibridación de sondas de ADN, la variabilidad de antígenos de superficie y los productos de excreción-secreción (Nash *et al.*, 1985; Nash y Keister, 1985). Apenas tres años más tarde, se definieron los grupos I, II, III y IV en función de la variación alozímica (Meloni *et al.*, 1988; Andrews *et al.*, 1989). Éstos fueron posteriormente reagrupados en los ensamblajes A (I y II) y B (III y IV), quedando definidos los subensamblajes AI, AII, BIII y BIV (Mayrhofer *et al.*, 1995). Casi coetáneamente, (Homan *et al.*, 1992), clasificaron aislados de *G. duodenalis* en dos grupos denominados “Polaco” y “Belga” utilizando alozimas, hibridación con sondas de ADN y RAPD (*Random Amplification of Polymorphic DNA*). En esta época ya se sugiere que, dados los niveles de divergencia entre ensamblajes, *G. duodenalis* podría ser un complejo de especies y no una sola.

Algunas de las distintas agrupaciones descritas en ese estrecho margen de tiempo demostraron ser equivalentes: por un lado el ensamblaje A y los grupos 1, 2 (AI y AII) y “Polaco”, y por otro el ensamblaje B y los grupos 3 y “Belga” (Monis *et al.*, 1996). La nomenclatura propuesta por Mayrhofer *et al.* (1995), prevaleció sobre las demás y es la que se mantiene hoy en día. Hasta la fecha se han descrito ocho ensamblajes: A, B (Mayrhofer *et al.*, 1995), C, D (Monis *et al.*, 1998), E, F, G (Monis *et al.*, 1999) y H (Lasek-Nesselquist *et al.*, 2010), de los cuales los ensamblajes A y B parasitan preferentemente humanos (Tabla 1.2), aunque también se han registrado esporádicamente algunos casos de aislados humanos pertenecientes a los ensamblajes C, D, E y F (Feng y Xiao, 2011).

Estos ensamblajes presentan valores de divergencia genética similares a los descritos entre distintas especies de protozoos como *Theileria parva* y *T. annulata*, *Leishmania major* y *L. infantum* o *Cryptosporidium hominis* y *C. parvum* (Jarlström-Hultqvist *et al.*, 2010). Además, están adaptados a distintos tipos de hospedadores, que pueden ir desde un amplio rango de mamíferos (como en los ensamblajes A y B), hasta prácticamente una única especie (como los ensamblajes F o G) (Ryan y Cacciò, 2013; Heyworth, 2016). Por ello en los últimos años varios autores han propuesto el reconocimiento del rango de especie para estos ensamblajes, recuperando algunos de los nombres específicos que quedaron en su momento bajo el paraguas del grupo morfológico *G. duodenalis* (Tabla 1.2) (Monis *et al.*, 2003;

Thompson y Monis, 2004; Cacciò y Ryan, 2008; Franzén *et al.*, 2009; Monis *et al.*, 2009; Jerlström-Hultqvist *et al.*, 2010; Xu *et al.*, 2012; Ryan y Cacciò, 2013).

Tabla 1.2: Ensamblajes de *G. duodenalis*, nombre específico propuesto y hospedadores.

Ensamblaje	Especie propuesta	Hospedadores
A	<i>G. duodenalis</i>	Humanos y otros primates, perros, gatos, ganado bovino y otros rumiantes, ovejas, cabras, cerdos, caballos, castores, cetáceos, pinnípedos, aves de corral, gaviotas
B	<i>G. enterica</i>	Humanos y otros primates, perros, caballos, castores, ganado bovino y otros rumiantes, ovejas, conejos, roedores, aves de corral, gaviotas, cetáceos pinnípedos
C/D	<i>G. canis</i>	Principalmente perros y otros cánidos, también ganado vacuno, cetáceos, canguros y cerdos
E	<i>G. bovis</i>	Principalmente ganado (bovino, ovino y porcino), también caballos, cabras, gatos, zorros
F	<i>G. cati</i>	Principalmente gatos, también cerdos y cetáceos
G	<i>G. simondi</i>	Ratas y ratones
H	–	Pinnípedos, gaviotas

El estudio de las relaciones filogenéticas reveló que los ensamblajes A y B no están estrechamente relacionados sino que se parecen más a otros ensamblajes que entre sí. Así, el análisis de alozimas reveló una mayor similitud del ensamblaje A con los ensamblajes E y F, que con el ensamblaje B (Figura 1.3) (Monis *et al.*, 2003). Esto concuerda con el análisis filogenético de secuencias de ADN, que agrupó por un lado a los ensamblajes A, E, F, y por otro a los ensamblajes B, C y D (Lasek-Nesselquist *et al.*, 2010; Feng y Xiao, 2011; Fernández-Álvarez *et al.*, 2014).

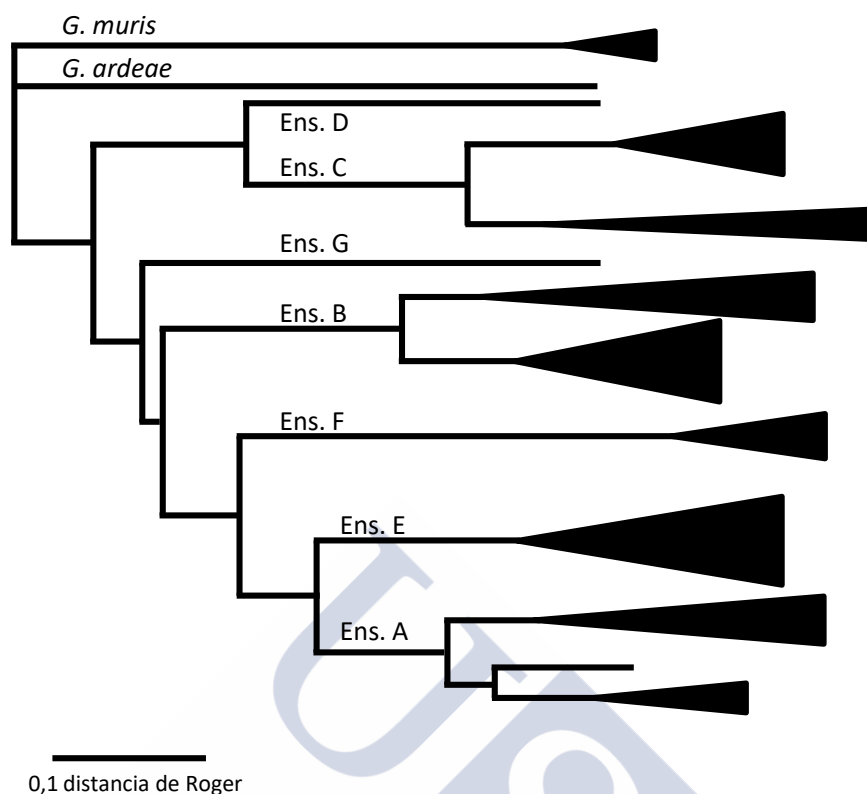


Figura 1.3: Dendrograma de las relaciones genéticas entre los ensamblajes de *G. duodenalis*. Adaptado de Monis *et al.* (2003). Las distancias genéticas fueron calculadas a partir de datos de electroforesis de enzimas y las relaciones genéticas fueron inferidas con el método Neighbor-Joining.

1.3.4 Variabilidad dentro de ensamblajes

La utilización de secuencias de ADN para la caracterización genética de los aislados supuso un salto cualitativo en la descripción de la variabilidad genética de *G. duodenalis*. La utilización de alozimas y otros marcadores clásicos como los RFLPs (*Restriction Fragment Length Polymorphism*) permitió diferenciar entre ensamblajes e incluso subensamblajes (Andrews *et al.*, 1989; Mayrhofer *et al.*, 1995; Homan *et al.*, 1998; Yong *et al.*, 2002; Cacciò *et al.*, 2002; Amar *et al.*, 2003), pero su baja resolución (solo permiten detectar variantes genéticas que provoquen diferencias en la migración electroforética y que afecten a las dianas de restricción, respectivamente) impidió describir en mayor detalle la variabilidad dentro de los ensamblajes. La secuenciación Sanger solventó los problemas de resolución de las técnicas anteriores, puesto que detecta todas las mutaciones que existen en la región

analizada. La estrategia más habitual consiste en amplificar por PCR uno o varios marcadores genéticos en cada aislado y secuenciarlos directamente (e.g., Lalle *et al.*, 2005; Robertson *et al.*, 2006; Wielinga y Thompson, 2007; Cacciò *et al.*, 2008; Lebbad *et al.*, 2008; de Lucio *et al.*, 2015).

Para el genotipado de aislados de *Giardia* mediante secuenciación se han utilizado principalmente cuatro marcadores: *glutamato deshidrogenasa* (*gdh*), *beta-giardina* (*bg*), *triosa fosfato isomerasa* (*tpi*) y la subunidad pequeña del ARNr (*SSU rRNA*), aunque también se han estudiado en menor medida otros loci, tanto codificadores, como los genes que codifican *proteínas de superficie variante* (*VSP*), la *proteína ribosómica L7a* (*rpL7a*), la *proteína mlh1* (*mlh1*), la *ferredoxina* [2Fe-2S] (*fd*), el *factor de elongación 1 alfa* (*ef1*), las *histonas H2B* y *H4*, la *chaperonina 60*, o la *fase de lectura abierta C4*, (*open reading frame C4*) como no codificadores, por ejemplo, la fracción comprendida entre la subunidad pequeña y el segundo espaciador interno del ARN ribosómico (*ITS1-5.8S rRNA-ITS2*) (revisado recientemente en Ryan y Cacciò, 2013).

El análisis de estos marcadores permitió obtener estimas más precisas de los niveles de variabilidad dentro de los ensamblajes. Las secuencias de aislados de origen humano confirmaron la existencia de dos grupos dentro del ensamblaje A que se corresponden con los subensamblajes AI y AII. En cambio, no se encontraron evidencias que justificasen la división del ensamblaje B en los subensamblajes BIII y BIV.

La asignación de los aislados a los ensamblajes y subensamblajes se suele realizar por homología de secuencia con los aislados que fueron utilizados para definirlos (Mayrhofer *et al.*, 1995). Sin embargo, a medida que aumentó el número de muestras analizadas, aparecieron secuencias de asignación dudosa, algunas de las cuales presentaban dobles picos en las posiciones diagnóstico (Lebbad *et al.*, 2011), incluso entre las obtenidas de un único quiste (Ankarklev *et al.*, 2012). En una revisión de la diversidad dentro de los ensamblajes se utilizaron los términos “BIII-like”, “BIV-like” e incluso “B-central” para describir algunas de las secuencias de *gdh* y *tpi*, dada la dificultad para asignarlas a uno u otro subensamblaje, mientras que en *bg* se evitó establecer cualquier tipo de división (Wielinga y Thompson, 2007).

La nomenclatura utilizada para referirse a las variantes no está estandarizada, ya que distintos autores utilizaron los términos subtipo (Sulaiman *et al.*, 2003; Gelanew *et al.*, 2007; Cacciò *et al.*, 2008; Lebbad *et al.*, 2011), genotipo (Sprong *et al.*, 2009), subgenotipo (Lalle *et al.*, 2005; Hussein *et al.*, 2009) o variante intra-genotípica (Wielinga y Thompson, 2007) indistintamente.

Tampoco hay un sistema estándar para nombrar los distintos alelos. El primer análisis de diversidad nucleotídica en el *locus bg* utilizó A1, A2 y A3 para denominar a las secuencias distintas detectadas dentro del ensamblaje A, y B1, B2, B3 y B4 para las del ensamblaje B (Cacciò *et al.*, 2002), sin hacer referencia a los subensamblajes. Esta práctica se extendió a otros estudios y loci, llegando a describirse un subtipo B19 en *bg* (Gelanew *et al.*, 2007) o un subtipo B0 en *gdh*. Otros autores utilizaron S1, S2, S3 (de subtipo) para referirse a distintas secuencias de *tpi* (Sulaiman *et al.*, 2003). Y otros mantuvieron la nomenclatura de los subensamblajes añadiéndoles sufijos que designasen el genotipo: BIV-var1, BIV-var2 y BIV-var3 (Geurden *et al.*, 2009). El uso de distintas nomenclaturas genera importantes problemas a la hora de comparar resultados o asociar genotipos de distintos loci, entre otros.

En el caso de *Giardia* el uso de una nomenclatura común no resolvió estas dificultades, sino que añadió otras. Los primeros estudios de diversidad analizaron un único locus y nombraron nuevos genotipos sin tener en cuenta los genotipos que se habían descrito utilizando otros aislados y otros loci. Por ello, un aislado genotipado como A1 en *bg* no tiene por qué estar relacionado con un aislado genotipado como A1 en *gdh*, lo que conlleva que identificar un aislado como genotipo A1 carezca de significado si no se especifica el locus al que se hace referencia. Por si la confusión no fuese ya suficiente, en algunos casos, las variantes similares a “genotipos” previamente descritos empezaron a designarse añadiendo números detrás de la denominación original, como B1-1 hasta B1-6 (Lebbad *et al.*, 2011).

Recientemente se ha señalado la necesidad de establecer una nomenclatura consenso y se ha propuesto la utilización del término genotipo para la descripción de nuevas variantes genéticas y genotipo multilocus en los casos en los que se dispone de la información de varios loci en un mismo aislado (Ryan y Cacciò, 2013). Así mismo se ha propuesto que para nombrar nuevos genotipos multilocus se debería utilizar el nombre del subensamblaje seguido de un guión y un número arábigo que identifique el genotipo: AII-1.

1.4 GENOMA

Giardia presenta un genoma haploide de aproximadamente 12 Mb distribuidas en cinco cromosomas (Upcroft *et al.*, 2010; Shen *et al.*, 2011), que resulta pequeño en comparación con el de otros eucariotas secuenciados hasta la fecha (Figura 1.4). Se trata de un genoma muy compacto, en el que entre un 75% y un 90% corresponde a secuencias codificantes de proteínas (Adam *et al.*, 2013), con regiones intergénicas cortas, y en el que los promotores y las regiones no traducidas (UTR) se encuentran reducidas al mínimo (Morrison *et al.*, 2007).

Inicialmente se identificaron solo cuatro intrones (*cis*-spliced) en todo el genoma, pero recientemente se ha descrito una nueva forma de *trans*-splicing mediada por el espliceosoma que es esencial para la maduración del ARN mensajero de al menos dos genes importantes. Algunos genes codificadores de proteínas parecen estar divididos en varios fragmentos esparcidos por el genoma de modo que cada fragmento se transcribe independientemente, generando varios pre-ARNs, que posteriormente interaccionan para ser combinados (*trans*-spliced) en el ARN maduro (Kamikawa *et al.*, 2011).

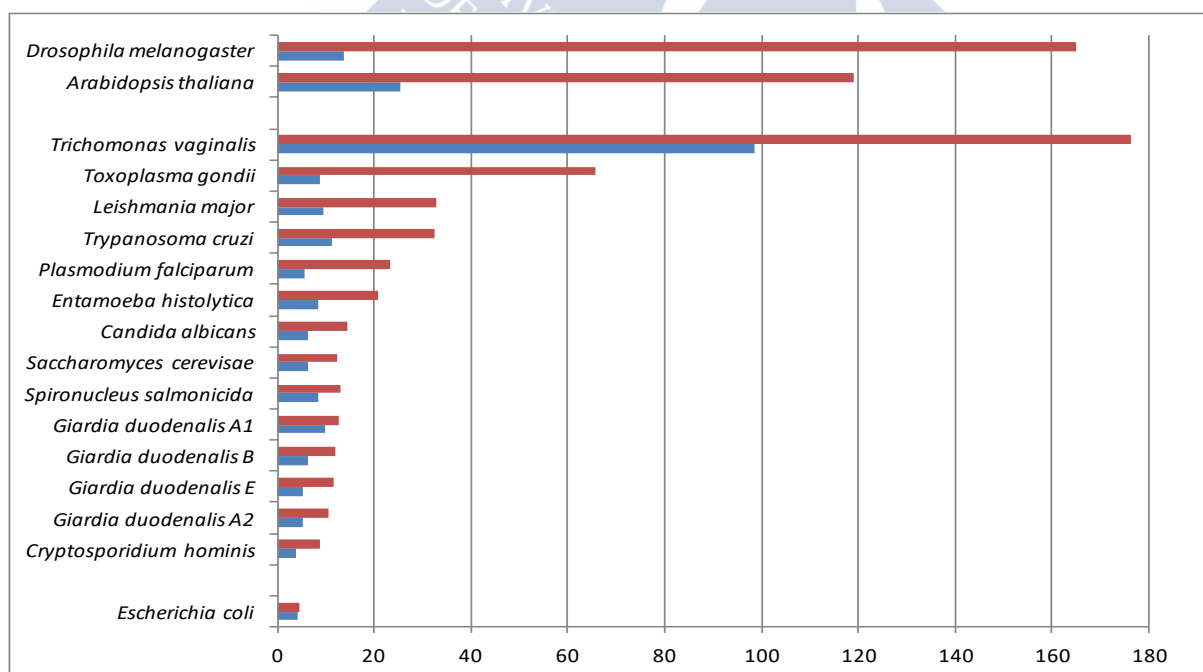


Figura 1.4: Comparación del tamaño del genoma de *Giardia* respecto a otros organismos. En rojo el tamaño del genoma expresado en Mb, y en azul el número de genes anotados en miles.

Hasta la fecha, en la base de datos GiardiaDB hay depositadas las secuencias de cinco genomas completos obtenidos mediante la secuenciación de cultivos axénicos de tres aislados de origen humano: WB subensamblaje A1 (Morrison *et al.*, 2007), DH subensamblaje A2, (Adam *et al.*, 2013) y GS ensamblaje B, secuenciado en dos ocasiones de forma independiente (Franzén *et al.*, 2009; Adam *et al.*, 2013); y uno de origen porcino: P15 ensamblaje E, (Jerlström-Hultqvist *et al.*, 2010) (Tabla 1.3). Aparte de estos cinco genomas de referencia, se han secuenciado dos genomas del subensamblaje A2 (AS175 y AS98; Ankarklev *et al.*, 2015) y un genoma del ensamblaje B (BAH15c1; Wielinga *et al.*, 2015) a partir de cultivos axénicos y dos genomas del ensamblaje B y uno del ensamblaje A2 obtenidos de la secuenciación de quistes purificados y concentrados directamente a partir de muestras clínicas (Hanevik *et al.*, 2015). Sin embargo de estos últimos solo están accesibles en GenBank las secuencias de los aislados AS175 y BAH15c1.

Tabla 1.3: Características de los genomas de *G. duodenalis* publicados.

Ensamblaje	Aislado/ Cultivo	Tamaño (Mb)	GC%	Genes	Proteínas	ASH	Scaffolds	Referencia
AI	WB C6	12,8	49,2	9747	9667	<0,01%	92	(Morrison <i>et al.</i> , 2007; Upcroft <i>et al.</i> , 2010)
B	GS/M	11,0	47,3	4550	4470	0,53%	2931	(Franzén <i>et al.</i> , 2009)
E	P15	11,5	47,2	5150	5007	<0,01%	820	(Jerlström-Hultqvist <i>et al.</i> , 2010)
AII	DH	10,7	49,0	5204	5147	0,04%	239	(Adam <i>et al.</i> , 2013)
B	GS	12,0	48,2	6165	6098	0,43%	543	(Adam <i>et al.</i> , 2013)
B	BAH15c1	10,4	47,0	4991	4910	-	508	(Wielinga <i>et al.</i> , 2015)
AII	AS175 *	10,5	48,9	-	-	-	1139	(Ankarklev <i>et al.</i> , 2015)

Nota: *Genoma parcial obtenido por secuenciación de ARN.

El genoma con la anotación más completa es el del aislado WB, que fue el primero en publicarse y que está ensamblado a nivel cromosómico (Upcroft *et al.*, 2010). En este genoma se han anotado hasta 9667 secuencias codificadoras, a pesar de que en un primer ensamblado

solo se identificaron 6470 pautas de lectura abiertas, un número más próximo al observado en el resto de genomas de *Giardia* (Tabla 1.3). La identidad nucleotídica promedio entre A y E es del 87% (Jerlström-Hultqvist *et al.*, 2010) y entre éstos y B, del 77% (Franzén *et al.*, 2009), confirmando la separación entre dichos ensamblajes observada en análisis filogenéticos previos (Monis *et al.*, 1999).

Los ensamblajes presentan grandes diferencias con respecto a su nivel de heterozigosidad (allelic sequence heterozygosity, ASH; Baruch *et al.*, 1996), entendido como el número de posiciones en las que se observan dos o más variantes alélicas respecto a la longitud total del genoma. La heterozigosidad en los genomas del ensamblaje B es más de 50 veces mayor que la observada en los ensamblajes A y E (Franzén *et al.*, 2009) (Tabla 1.3).

1.5 EPIDEMIOLOGÍA MOLECULAR

En parasitología, la epidemiología molecular y la ecología molecular se podrían definir como la aplicación de técnicas genético-moleculares al estudio de las dinámicas poblacionales de la enfermedad vistas desde el punto de vista del hospedador y del parásito, respectivamente. No se puede llegar a entender la epidemiología de la giardiasis sin conocer la ecología de *G. duodenalis*, empezando por su ciclo vital, y para ello es fundamental el estudio de la genética y la evolución de sus poblaciones.

Entre las principales aplicaciones de la epidemiología molecular se encuentran: i) el desarrollo de herramientas que permitan identificar los patógenos de forma específica y sensible (tanto la especie como otras unidades taxonómicas subespecíficas), ii) la resolución de conflictos taxonómicos (sobre todo a nivel subespecífico), iii) el estudio del ciclo de transmisión del parásito y sus implicaciones epidemiológicas (*e.g.*, existencia de zoonosis), iv) el análisis de su interacción con otros patógenos u otros linajes de la misma especie (frecuencia e implicaciones clínicas de las infecciones mixtas), v) el establecimiento de asociaciones genotipo-fenotipo de caracteres de relevancia clínica (*e.g.*, respuesta a fármacos, virulencia, infectividad,...) (Lymbery y Thompson, 2012). El conocimiento de los patrones de diversidad genética de las poblaciones del parásito es fundamental para establecer un marco de trabajo adecuado que permita resolver estas cuestiones.

El genotipado de un número cada vez mayor de aislados, normalmente mediante secuenciación directa de *bg*, *gdh* y/o *tpi*, en poblaciones de todo el mundo ha revelado un escenario realmente complejo (Lalle *et al.*, 2005; Gelanew *et al.*, 2007; Robertson *et al.*, 2007; Teodorovic *et al.*, 2007; Lebbad *et al.*, 2008; Cacciò *et al.*, 2008; Hussein *et al.*, 2009; Lasek-Nesselquist *et al.*, 2009; Kosuwin *et al.*, 2010; Lebbad *et al.*, 2011; Siripattanapipong *et al.*, 2011; Ankarklev *et al.*, 2012; Huey *et al.*, 2013; Durigan *et al.*, 2014; Mateo *et al.*, 2014; Choy *et al.*, 2015; de Lucio *et al.*, 2015; Minetti *et al.*, 2015; revisado en Cacciò y Sprong, 2010; Feng y Xiao, 2011; Ryan y Cacciò, 2013; Thompson y Ash, 2016). La existencia de infecciones mixtas de distintos ensamblajes o de distintos linajes de un mismo ensamblaje ha complicado la interpretación de los resultados de tal modo que ni tan siquiera hay consenso acerca de la existencia de recombinación genética entre ensamblajes (Teodorovic *et al.*, 2007; Lasek-Nesselquist *et al.*, 2009; Sprong *et al.*, 2009; Almeida *et al.*, 2010; Xu *et al.*, 2012; revisado en Cacciò y Sprong, 2010).

1.5.1 Determinación de ensamblaje

Los aislados de origen humano generalmente contienen giardias de los ensamblajes A y B (ver apartado 1.3.3), aunque ocasionalmente también se han detectado otros ensamblajes (<1% del total en una encuesta de más de 1.500 aislados de toda Europa, Sprong *et al.*, 2009) (Feng y Xiao, 2011; Choy *et al.*, 2015).

Por regla general las infecciones causadas por el ensamblaje B (aprox. 60%) son más frecuentes que las causadas por el ensamblaje A (aprox. 40%) (Thompson y Monis, 2012), con una distribución similar en países industrializados y no industrializados, aunque los datos varían considerablemente entre países y estudios (revisado en Feng y Xiao, 2011). En numerosas ocasiones se detectaron infecciones mixtas, con los ensamblajes A y B (revisado en Feng y Xiao, 2011; Ryan y Cacciò, 2013). Este fenómeno es más frecuente en los países en vías de desarrollo (8% vs. 2% en países desarrollados; Thompson y Monis, 2012), pero los datos también varían ampliamente en función de la región geográfica y de la metodología empleada para el genotipado, llegando a alcanzar el 25% en un estudio que utilizó cebadores específicos para detectar cada ensamblaje por separado en Etiopía (Gelanew *et al.*, 2007).

En los estudios en los que se secuenciaron varios loci no es extraño detectar incongruencias entre los ensamblajes detectados en los distintos loci (Read *et al.*, 2004; Cacciò *et al.*, 2008; Sprong *et al.*, 2009). El análisis de las secuencias depositadas en una base de datos Europea con 2476 aislados de origen humano, animal y ambiental (ZOOPNET, Sprong *et al.*, 2009), reveló inconsistencias entre loci en un 13% de los aislados (un 12% en los de origen humano). Este patrón se reprodujo a nivel de subensamblaje: los aislados con genotipos multilocus inconsistentes para los subensamblajes AI y AII, representaron un porcentaje similar a los que fueron designados como AI en todos los loci (5%) y los inconsistentes para BIII y BIV alcanzaron el 30% (Sprong *et al.*, 2009). Estos resultados pueden tener dos causas: la recombinación entre ensamblajes y subensamblajes o la amplificación preferencial de distintos ensamblajes o subensamblajes en los distintos loci en una infección mixta.

En vista de la variabilidad de los síntomas de la giardiasis, una de las primeras aplicaciones del genotipado de aislados fue el estudio de la asociación entre los ensamblajes y la sintomatología. Los estudios realizados hasta la fecha ofrecieron resultados contradictorios. Varios autores señalaron una posible asociación del ensamblaje A con una sintomatología más severa y del ensamblaje B con infecciones asintomáticas en estudios realizados en distintas poblaciones (Read *et al.*, 2002; Aydin *et al.*, 2004; Haque *et al.*, 2005; Sahagún *et al.*, 2008), mientras que otros sugirieron todo lo contrario (Homan y Mank, 2001; Gelanew *et al.*, 2007; Mohammed Mahdy *et al.*, 2009; Puebla *et al.*, 2014; Minetti *et al.*, 2015).

La influencia de coinfecciones con otros parásitos podría explicar en parte los resultados contradictorios, aunque la utilización de categorías genéticas tan amplias como los ensamblajes, los problemas que suponen las infecciones mixtas A+B y la obtención de genotipos ambiguos, unidos a la dificultad para categorizar la sintomatología (tipo de heces; diarrea aguda, crónica, persistente;...) hacen que establecer una correlación genotipo-fenotipo sea extremadamente complejo. El futuro éxito de los estudios de asociación reside en la capacidad para definir correctamente tanto el genotipo de los aislados como el fenotipo clínico del paciente.

1.5.2 Diversidad genética

Existe un amplio consenso acerca de la mayor variabilidad genética del ensamblaje B respecto al ensamblaje A en los aislados de origen humano (Plutzer *et al.*, 2010; Wielinga *et al.*, 2015). Sin embargo, el uso extendido de la secuenciación directa de los productos de amplificación ha impedido una correcta estima de la diversidad genética en la mayor parte de los casos.

La secuenciación directa por Sanger permite obtener una única secuencia por locus para cada muestra analizada, que se suele identificar como el genotipo del aislado. En numerosos estudios que utilizaron esta técnica la aparición de dobles picos en los electroferogramas reveló la existencia de diversidad dentro de los aislados (Lalle *et al.*, 2005; Robertson *et al.*, 2007; Cacciò *et al.*, 2008; Lebbad *et al.*, 2008; Lebbad *et al.*, 2011; Huey *et al.*, 2013; Durigan *et al.*, 2014; Choy *et al.*, 2015; de Lucio *et al.*, 2015; Minetti *et al.*, 2015). Los dobles picos son señales superpuestas en posiciones polimórficas, e indican que hay más de un alelo presente en la muestra, pero no permiten establecer la fase haploide en que se encuentran las distintas variantes. Además, la señal de las variantes que se encuentran a alta frecuencia puede enmascarar a las que son menos comunes, lo que deriva en una subestima de los verdaderos niveles de polimorfismo en las muestras.

La presencia de dobles picos es mayor en aislados del ensamblaje B que en los del A, en concordancia con la mayor variabilidad observada en este ensamblaje (21% de 1151 secuencias del ensamblaje B y 5% de 1290 del A; Sprong *et al.*, 2009). Este hecho ha limitado la determinación de genotipos multilocus en el ensamblaje B, porque la metodología empleada hace imposible diferenciar entre la presencia de varios genotipos homocigotos o un único genotipo heterocigoto (Lebbad *et al.*, 2011; Minetti *et al.*, 2015).

Las secuencias con posiciones ambiguas (dobles picos) tienen poco valor para el estudio de la genética de poblaciones, ya que no permiten identificar los alelos subyacentes, que son los datos genéticos de relevancia evolutiva y ecológica. Estos contratiempos se solventan mediante la clonación de los productos de PCR previa a la secuenciación, que permite obtener las secuencias de los diversos haplotipos presentes en una infección mixta. Sin embargo este proceso consume tiempo y recursos, ya que para detectar los alelos presentes en un aislado es necesario secuenciar un número elevado de clones.

Solo un pequeño número de estudios han utilizado hasta el momento la clonación de los productos de PCR para la descripción de la diversidad genética en aislados de *G. duodenalis* de origen humano: Teodorovic *et al.* (2007), Lasek-Nesselquist *et al.* (2009), Hussein *et al.* (2009), Kosuwin *et al.* (2010) y Siripattanapipong *et al.* (2011). Los dos primeros utilizaron una estrategia multilocus en cultivos axénicos, mientras que los tres últimos clonaron un único locus en aislados de pacientes: *gdh* (Hussein *et al.*, 2009; Siripattanapipong *et al.*, 2011) y *bg* (Kosuwin *et al.*, 2010). Tanto en los trabajos con cultivos axénicos como en el trabajo de Kosuwin *et al.* se secuenciaron clones en todos los aislados analizados, mientras que en los otros dos los productos de PCR fueron secuenciados directamente y clonados solo en los aislados en los que se detectaron dobles picos.

Estos estudios confirmaron las diferencias en los niveles de diversidad genética en las poblaciones naturales de los ensamblajes A y B, principalmente en las posiciones sinónimas (*gdh*: 0,19% vs. 3,90%; *bg*: 1,15% vs. 2,17%, para A y B respectivamente). La variación en las posiciones no sinónimas fue mucho menor en ambos ensamblajes (*gdh*: 0% vs. 0,10%; *bg*: 0,42% vs. 0,26%, para A y B respectivamente). Dentro de los aislados también se detectó mayor variabilidad en el ensamblaje B (número de aislados en los que se detectó más de un haplotipo, *gdh*: 0/20 vs. 9/22; *bg*: 1/12 vs. 14/18; para A y B respectivamente), aunque en este caso no se cuantificó la diversidad genética (datos extraídos de Siripattanapipong *et al.*, 2011; Kosuwin *et al.*, 2010). La presencia de múltiples secuencias dentro de un hospedador puede deberse a una infección mixta o bien a diferencias entre los alelos de cada individuo (ASH).

1.5.3 Potencial zoonótico

El potencial zoonótico de *G. duodenalis* es una de las cuestiones que ha recibido mayor atención en el estudio de este parásito, dado el amplio rango de especies animales en las que se han detectado los ensamblajes que parasitan principalmente humanos: A y B (revisado en Feng y Xiao, 2011; Ryan y Cacciò, 2013) (Tabla 1.2). Por ejemplo, tan solo en la península Ibérica, además de en humanos, los ensamblajes A y B han sido detectados en perros (Dado *et al.*, 2012; Gil *et al.*, 2017), en ganado vacuno (Mendonca *et al.*, 2007), en ovejas (Gómez-Muñoz *et al.*, 2009; Gómez-Muñoz *et al.*, 2012), en numerosas especies de aves salvajes (Reboredo-Fernández *et al.*, 2015; Cano *et al.*, 2016), en cetáceos (Reboredo-Fernández *et al.*, 2014; Reboredo-Fernández *et al.*, 2015), y otros mamíferos salvajes (García-Presedo *et al.*,

2013). Esto abre la posibilidad de que existan vías de transmisión entre animales y humanos. A nivel de subensamblaje se mantiene el potencial zoonótico, ya que AI, AII, BIII y BIV han sido detectados tanto en humanos como en animales (revisado en Feng y Xiao, 2011; Ryan y Cacciò, 2013).

Sin embargo, es posible que dentro de cada ensamblaje o subensamblaje, humanos y animales sean parasitados por distintos genotipos adaptados a cada hospedador. Por ejemplo, el subensamblaje AIII (Cacciò y Ryan, 2008), ha sido detectado principalmente en ungulados salvajes y en menor medida en ganado y gatos, pero no en humanos. Las evidencias de zoonosis provienen de la identificación de exactamente el mismo genotipo de *G. duodenalis* en aislados de origen humano y animal. Desafortunadamente, los problemas en el genotipado (secuencias ambiguas e inconsistencia entre loci, descritos en el apartado anterior), interfieren en la identificación inequívoca de los aislados y su comparación. Este hecho ha impedido determinar de forma precisa el papel que juega la transmisión entre humanos y animales en la epidemiología de la giardiasis (Feng y Xiao, 2011; Ryan y Cacciò, 2013).

Otra de las limitaciones en la comparación de aislados es el uso de un número reducido de marcadores en el genotipado. Un aislado de origen humano y otro de origen animal pueden ser considerados zoonóticos cuando se analiza un único locus, pero mostrar un genotipo específico de un hospedador cuando se incluyen más loci en el análisis. Por ejemplo, en el análisis de las secuencias depositadas en ZOOPNET se observó que alrededor de un 15% de las secuencias de los loci *gdh*, *bg* y *tpi* se detectaron tanto en aislados de origen humano como animal. Sin embargo, cuando se analizaron los datos de estos loci asociados en pares, tan solo 8 de los 102 genotipos multilocus de A fueron detectados tanto en humanos como en animales, y solo 1 de los 159 del ensamblaje B. Si se tienen en cuenta los genotipos multilocus de los tres loci, tan solo 2 genotipos del ensamblaje A tendrían potencial zoonótico y ninguno del ensamblaje B (Sprong *et al.*, 2009).

Estos datos reflejan la dificultad para determinar el potencial zoonótico de *G. duodenalis* debido a la escasez de genotipos multilocus no ambiguos disponibles, y pone de manifiesto la necesidad de establecer protocolos de genotipado multilocus estandarizables para obtener datos comparables entre estudios que permitan esclarecer estas cuestiones.

1.5.4 Recombinación

En las últimas dos décadas se han ido recopilando evidencias indirectas de recombinación e intercambio genético en *G. duodenalis* desde distintos campos de la biología como la genética de poblaciones, la genómica comparada o la citogenética, aunque no está claro si se produce mediante un ciclo sexual o parasexual (revisado en Birky, 2005; Birky, 2010; Andersson, 2012) ni si ocurre entre ensamblajes o solo entre individuos de un mismo ensamblaje.

La escasa variabilidad genética observada dentro de algunos aislados tanto en la secuenciación de los genomas de referencia (Morrison *et al.*, 2007; Franzén *et al.*, 2009; Jerlström-Hultqvist *et al.*, 2010; Adam *et al.*, 2013) como en el genotipado de cultivos axénicos o aislados de pacientes (e.g., Lasek-Nesselquist *et al.*, 2009; Sprong *et al.*, 2009; Kosuwin *et al.*, 2010; Siripattanapipong *et al.*, 2011; Huey *et al.*, 2013; de Lucio *et al.*, 2015; Minetti *et al.*, 2015) se considera una evidencia a favor de la existencia de recombinación en este organismo (Birky, 2005). El efecto Meselson predice que, bajo reproducción asexual estricta, los cromosomas homólogos de especies diploides acumularán mutaciones con el paso de las generaciones hasta ser más distintos entre sí que respecto a sus respectivos cromosomas ortólogos de otros linajes cercanos (Mark Welch y Meselson, 2000). Este fenómeno debería ser aún más marcado en *Giardia* ya que cada trofozoíto tiene dos núcleos diploides ($2 \times 2n$), por lo que cabría esperar acumulación de diferencias entre los cromosomas de cada núcleo y entre los distintos núcleos. Por esta razón, la falta de variabilidad observada en numerosos aislados implica la existencia de un mecanismo de recombinación capaz de erosionar la diversidad que se genera por mutación.

En el genoma de *Giardia* se identificó una batería de 13 genes funcionales homólogos a los implicados en la meiosis en otros eucariotas, cinco de los cuales parecen específicos de este proceso, ya que no se les conoce otra función (Ramesh *et al.*, 2005). Aunque la presencia de estos genes no implica necesariamente que sigan cumpliendo esta función en *Giardia*, donde nunca se ha observado una fase meiótica, confieren verosimilitud a la posibilidad de que *Giardia* tenga la capacidad de completar una fase sexual o parasexual similar a la meiosis.

Otra de las evidencias de recombinación es la diplomixis. Como se explica en el apartado 1.2.3, Poxleitner *et al.* (2008) describieron la fusión de las envueltas nucleares durante la enquistación y demostraron la posibilidad de intercambio de material genético entre núcleos en un proceso que definieron como parasexual. Además, los autores de este trabajo estudiaron la expresión de los cinco genes homólogos específicos de la meiosis durante el ciclo vital de *G. duodenalis* y demostraron que tres de ellos se expresan únicamente en el quiste. Esto les llevó a proponer que estos genes estarían implicados en un proceso de recombinación mitótica entre cromosomas homólogos en los núcleos fusionados, que podría explicar en parte la reducción de las diferencias esperadas entre los cromosomas de los dos núcleos.

Por último, (Cooper *et al.*, 2007) demostraron la existencia de intercambio genético entre individuos en el subensamblaje AII. Bajo la hipótesis de reproducción asexual, se espera que los distintos loci se encuentren totalmente ligados y segreguen conjuntamente generación tras generación. De este modo la topología de un árbol filogenético de un conjunto de aislados debería ser la misma independientemente de la región del genoma que se haya utilizado (Posada *et al.*, 2002). Sin embargo, el análisis de largas regiones de los cromosomas 3, 4 y 5 (6 Kb, 1.2 Kb y 9.6 Kb, respectivamente) en cinco aislados del subensamblaje AII, reveló distintas historias evolutivas en distintos cromosomas, lo que sugiere la existencia de un proceso de segregación cromosómica. Además, al alinear las secuencias obtenidas detectaron puntos de recombinación en todas las regiones cromosómicas, lo que implica que también existe recombinación dentro de los cromosomas, que podría ser de origen mitótico (conversión génica) o meiótico (Cooper *et al.*, 2007).

Algunos autores utilizaron las inconsistencias en la determinación de los ensamblajes de los aislados (ver apartado 1.5.2) y la aparición de secuencias quiméricas para proponer que la recombinación se produce incluso entre individuos de distintos ensamblajes (Teodorovic *et al.*, 2007; Lasek-Nesselquist *et al.*, 2009); sin embargo, el análisis de los genomas de A, B y E no detectó evidencias de recombinación (Xu *et al.*, 2012).

La existencia de recombinación tendría importantes implicaciones para la taxonomía de estos parásitos, ya que permitiría aplicar el concepto biológico de especie, de forma que aquellos taxones que intercambiasen material genético pertenecerían a la misma especie. En cuanto a la epidemiología, la recombinación afectaría principalmente a la identificación de los aislados (el genotipado) y por tanto al estudio del potencial zoonótico y la asociación de los

genotipos con el fenotipo clínico. Si los loci segregan de forma independiente, las regiones utilizadas para la identificación del genotipo de los aislados no serían representativas del genoma completo y un aislado podría ser identificado como potencialmente zoonótico con un marcador y como específico de un hospedador con otro, o asociado con una sintomatología grave o leve en función del marcador estudiado (Cacciò y Sprong, 2010).

1.5.5 Análisis de quistes individuales

Varios autores han señalado en los últimos años la necesidad de analizar quistes individuales para resolver el origen de la diversidad genética dentro de las infecciones (ASH vs. infecciones mixtas) y confirmar las señales de recombinación detectadas en los genotipos multilocus (Cacciò *et al.*, 2008; Feng y Xiao, 2011; Thompson y Monis, 2011; Takumi *et al.*, 2012).

Miller y Sterling (2007) demostraron la viabilidad de la amplificación por PCR a partir de quistes individuales de *Giardia*. Utilizando un protocolo que consistía en el aislamiento de quistes mediante micromanipulación y la amplificación de una región conservada de *tpi* mediante PCR anidada, consiguieron una tasa de amplificación del 80%. Otros autores han publicado distintos ensayos de PCR y qPCR con sensibilidad suficiente para detectar un único quiste, aunque en estos casos la estima de la sensibilidad del método se realizó mediante diluciones seriadas de cultivos de trofozoítos o ADN (Amar *et al.*, 2003; Almeida *et al.*, 2010).

Hasta la fecha existen dos ensayos en los que se ha llevado a cabo el genotipado de quistes individuales: (Ankarklev *et al.*, 2012; Aguiar *et al.*, 2015; Aguiar *et al.*, 2016). En ambos casos los quistes fueron aislados mediante micromanipulación y transferidos a un tubo de PCR donde se llevó a cabo la extracción de ADN. En el primero de ellos se secuenciaron regiones de *bg* o *tpi* (se utilizaron por separado en distintos quistes) y la presencia de dobles picos permitió a los autores verificar la existencia de ASH en un elevado número de quistes. Se detectó variación en posiciones que habían sido anteriormente utilizadas para diferenciar subensamblajes. En el segundo trabajo se amplificaron cuatro loci en multiplex y a continuación se reamplificó cada uno de ellos por separado utilizando PCRs anidadas. Solo uno de los diez quistes analizados (todos obtenidos de una misma muestra) mostró

heterocigosidad. La clonación de los productos de PCR dio como resultado un número inesperadamente alto de haplotipos (9+/locus) pertenecientes a los ensamblajes A y B. Estos resultados son difícilmente compatibles con la dotación genética y el ciclo biológico del parásito.





2. Objetivos



2 Objetivos

Objetivo 1: Describir los patrones de diversidad genética en una población natural de *G. duodenalis*.

Objetivo secundario 1.1: Cuantificar los niveles de variación de los ensamblajes A y B en aislados de pacientes del Área sanitaria de Santiago de Compostela.

Diseño experimental: Amplificación, clonación y secuenciación de tres loci nucleares. Para que sean representativos del genoma en su conjunto, se utilizan genes de copia única, no ligados y no sometidos a selección natural. La clonación de los productos de PCR permite aislar los haplotipos presentes en cada aislado de forma inequívoca. Así mismo el análisis de varios clones por aislado permite obtener estimas de la diversidad genética de la subpoblación infectante en cada paciente.

Objetivo secundario 1.2: Analizar la estructura genética de la población.

Diseño experimental: Análisis molecular de la varianza genética en los aislados, para determinar la contribución relativa de las diferencias entre y dentro de los aislados a la variación total.

Objetivo 2. Cuantificar la variación genética en quistes individuales de *G. duodenalis*.

Objetivo secundario 2.1: Determinar el origen de la heterogeneidad genética de las infecciones de *G. duodenalis*.

Diseño: Identificación de los genotipos multilocus (MLG, de las siglas en inglés *multilocus genotypes*) de quistes individuales de cinco aislados del ensamblaje B. Etapas: i) aislamiento de los quistes por micromanipulación, ii) amplificación mediante PCR directa multiplex de siete loci nucleares de copia única, no ligados y no sometidos a selección natural y iii) identificación de los alelos presentes en cada quiste mediante secuenciación masiva en paralelo (MPS, de las siglas en inglés *massive parallel sequencing*) de los amplicones. Las lecturas producidas mediante MPS son clonales y por tanto permiten aislar los haplotipos presentes en cada quiste y determinar de forma inequívoca las diferencias entre los haplotipos presentes en cada individuo.

Objetivo secundario 2.2: Identificar evidencias de recombinación genética.

Diseño: Análisis de las relaciones filogenéticas de los alelos de quistes individuales en siete loci por separado. La obtención de genotipos multilocus a partir de un único individuo permite establecer una asociación genética entre los alelos de los distintos loci. En ausencia de recombinación todo el genoma se hereda en bloque y los árboles filogenéticos inferidos en cada locus deberían presentar la misma topología. La detección de distintas asociaciones alélicas a lo largo de los siete loci, evidenciaría la existencia de recombinación, en contra de la hipótesis de reproducción asexual.





3. Material y métodos



3 Material y Métodos

3.1 EXPERIMENTO I: ANÁLISIS DE LOS PATRONES DE DIVERSIDAD GENÉTICA EN AISLADOS DE LOS ENSAMBLAJES A Y B

3.1.1 Amplificación, clonación y secuenciación

3.1.1.1 Origen de las muestras y extracción de ADN

Las muestras utilizadas en este experimento pertenecen a una colección de heces de pacientes con enfermedad gastrointestinal recogidas entre los años 2000 y 2010 en el Complejo Hospitalario Universitario de Santiago de Compostela (CHUS). El diagnóstico de giardiasis se realizó mediante examen microscópico de las mismas. Aquellas positivas para *Giardia* se mantuvieron congeladas a -20°C hasta su posterior análisis.

La amplificación de ADN de muestras fecales es compleja debido a la presencia de inhibidores de la PCR (polisacáridos complejos, ácidos biliares, lípidos,...) (Schrader *et al.*, 2012) y ADN de otros organismos (humano, bacteriano, vírico o parasitario). Para minimizar el efecto de los inhibidores se utilizó un kit comercial de extracción y purificación de ADN específico para muestras de heces (QIAamp DNA Stool Mini Kit, QIAGEN), que elimina los inhibidores más importantes presentes en este tipo de muestras.

3.1.1.2 Selección de genes y diseño de cebadores

El locus *tpi* (Genbank: GL50803_93938), que presenta una divergencia del 18,7% entre los ensamblajes A y B (comparación de las secuencias genómicas de referencia), se utilizó para realizar un cribado que permitiese determinar el ensamblaje de las muestras. Para el análisis de polimorfismo de secuencia se seleccionaron tres loci de copia única en el genoma de *Giardia*: *gdh* (GL50803_21942), *bg* (GL50803_4812) y *caltractina* (*calt*: GL50803_104685). Los tres primeros han sido extensamente estudiados por diversos investigadores, lo que se refleja en el elevado número de secuencias disponibles en las bases de datos. *Calt*, sin embargo, es un nuevo marcador del que solo se conocían, al inicio de este

trabajo, las secuencias obtenidas durante la secuenciación del genoma completo de los ensamblajes A (aislado WB), B (GS) y E (P15) y más tarde del subensamblaje AII (DH) y otro aislado del ensamblaje B (BAH15c1). Los cebadores utilizados para la amplificación de los distintos loci se indican en la Tabla 3.1.

Tabla 3.1. Cebadores y reacción de PCR en la que fueron utilizados.

Locus	Cebador	Secuencia	Tª de anillamiento	Reacción
<i>tpi</i>	Tpi F	GCCTGCTCGTCGYCCCTTC	55	PCR externa conservado
	A_tpiR	AGCGCCACCGACAAGGAAGC	55	PCR externa ens. A
	B_tpiR	TTTGTCCAGAGCACGCTTCGCC	55	PCR externa ens. B
	A_tpiF2	GTACACCTGTCAACAGCCAT	55	PCR interna ens. A
	A_tpiR2	GGAAGCCGTCAATATTCGGA	55	PCR interna ens. A
	B_tpiF2	ACGTAGCGTCCATCGCCT	55	PCR interna ens. B
	B_tpiR2	GGTCTCGCCCATGAYTCTAC	55	PCR interna ens. B
<i>gdh</i>	Gdh F	TTCCGTRTYCAGTACAAYTC	55	PCR externa conservado
	Gdh R	ACCTCGTTCTGRGTRGCGCA	55	PCR externa conservado
	Gdh F2	CTCGGSCCCTACAAGGGTGG	55	PCR interna conservado
	Gdh R2	TCCCRRTKGAGTCYGAGAAGGTGA	55	PCR interna conservado
	Gdh A F	CCTCTGTCAATCTTTCGATTCTC	55	PCR interna ens. A
	Gdh A R	CTTTGCGCCRAGCTGAATGA	55	PCR interna ens. A
	Gdh B F	CGGCTTTGAGCAGATCCTGA	55	PCR interna ens. B
	Gdh B R	GACCTTCGCRCCGAGC	55	PCR interna ens. B
	Gdh B F1	CCRATGGGCGGTGGT	55	PCR externa ens. B
	Gdh B R1	CACYCCCAGGGYTTCTTGTTTC	55	PCR externa ens. B
	Gdh B F3	CGGCTCCGACTTCGATCCT	55	PCR interna ens. B
	Gdh B R3	GTYGAAGYCATCCTTATCGACG	55	PCR interna ens. B
<i>bg</i>	Bg 1 F	TGAAYCAGCGCGTCAGCAGG	55	PCR externa conservado
	Bg 1 R	GGAGATGCGGGCGAAGCCCT	55	PCR externa conservado
	Bg 1 F1	YCAGCGCGTCAGCAGGTTCC	55	PCR interna conservado
	Bg 1 R1	AGATGCGGGCGAAGCCCTCT	55	PCR interna conservado
	Bg B F	CGATGACCTCACYCGCAGC	55	PCR externa ens. B
	Bg B R	AGCTTCGTGTTTGYGAGGGCC	55	PCR externa ens. B
	Bg B F1	CATCGCGCACCTYGACAGA	55	PCR interna ens. B
<i>calt</i>	Calt F	CATATCAACGGGGAAGCCTC	50	PCR externa conservado
	Calt R	GAGAAAGCACTYGTGGACCT	50	PCR externa conservado
	Calt F1	TATCAACGGGGAAGCCTCG	50	PCR interna conservado
	Calt R1	TCATGGCCAARAGCTCCTC	50	PCR interna conservado
	Calt B F	GAAGCCTCGCAGAAARACG	55	PCR interna ens. B
	Calt B R	AATTCCTTGGCAACACGCCT	55	PCR interna ens. B

Para el diseño de los cebadores se utilizó la herramienta online Primer-BLAST (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>). Las secuencias de estos loci depositadas en GiardiaDB (<http://giardiadb.org/giardiadb>, (Aurrecoechea *et al.*, 2009) y GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) se alinearon de modo que se pudiesen identificar regiones conservadas y divergentes entre ensamblajes. En las regiones conservadas se diseñaron parejas de cebadores que permitiesen amplificar ambos ensamblajes siguiendo una estrategia de PCRs anidadas. Por su parte, las regiones divergentes se utilizaron para diseñar cebadores específicos para cada ensamblaje.

3.1.1.3 Determinación del ensamblaje y selección de muestras

La cantidad de ADN de *Giardia* en el extracto final es muy reducida en comparación con el ADN humano y de otros organismos (hay que tener en cuenta que los quistes presentan una pared celular que hay que romper para acceder al contenido nuclear y que la mayoría de células presentes en las heces no tienen). Para aumentar la especificidad y sensibilidad de las PCRs se utilizó un protocolo de amplificaciones anidadas. Esta técnica consiste en la realización de dos PCRs consecutivas donde la primera –externa– funciona como una reacción de enriquecimiento del ADN diana, y la segunda –interna– amplifica la región deseada.

Para la determinación de los ensamblajes presentes en las muestras se utilizó el sistema de PCR DreamTaq (Thermo Scientific). En las PCRs externas se utilizó el cebador conservado TpiF en combinación con los cebadores específicos para cada ensamblaje A_tpiR y B_tpiR, y en las PCRs internas se utilizaron los cebadores A_tpiF2/A_tpiR2 y B_tpiF2/B_tpiR2 para la amplificación de los ensamblajes A y B, respectivamente (Tabla 3.1). Las condiciones de amplificación se detallan en las Tablas 3.2 y 3.3.

Los datos de recogida de los aislados que se seleccionaron para el análisis de diversidad de los ensamblajes A y B y los números de acceso de las secuencias depositadas en el *European Nucleotide Archive* (ENA) se indican en la Tabla 3.4.

Tabla 3.2: Condiciones de las reacciones de PCR para la determinación del ensamblaje.

	PCR externa	PCR interna
Tampón de PCR DreamTaq 10x	1 µL	1 µL
dNTPs	200 µM	200 µM
Cebadores	0.6 µM <i>Forward</i> 0.6 µM <i>Reverse</i>	0.6 µM <i>Forward</i> 0.6 µM <i>Reverse</i>
DreamTaq ADN polimerasa	0.25 U	0.25 U
AND molde	1-2 µL	1 µL de producto de la PCR externa
H ₂ O	hasta 10 µL	hasta 10 µL
Volumen Total	10 µL	10 µL

Tabla 3.3: Condiciones de amplificación de las PCR anidadas.

	Tiempo	PCR externa		PCR interna	
		T ^a	Ciclos	T ^a	Ciclos
Desnaturalización inicial	2'	95°C	1	95°C	1
Desnaturalización	30''	95°C		95°C	
Anillamiento	30''	52°C	35	55°C	25
Extensión	1'	72°C		72°C	
Extensión final	5'	72°C	1	72°C	1

Tabla 3.4: Detalles de los aislados incluidos en el análisis.

Aislados	Genotipado <i>tpi</i>	Fecha de Recogida	Números de Acceso (ENA)		
			<i>gdh</i>	<i>bg</i>	<i>calt</i>
122	A+B	Febrero-2002	LT603736-43	LT603860-67	LT603967-79
147	A	Abril-2002	LT603744-59	LT603868-75	LT603980-87
152	A	Mayo-2002	LT603760-67	LT603876-83	LT603988-95
209	A	Diciembre-2002	LT603768-83	LT603884-91	LT603996-03
251	A	Julio-2003	LT603784-91	LT603892-98	LT604004-11
263	A	Julio-2003	LT603792-99	LT603899-06	LT604012-19
321	A+B	Octubre-2003	LT603800-19	LT603907-24	LT604020-32
407	B	Diciembre-2003	LT603820-27	LT603925-32	LT604033-39
704	B	Noviembre-2005	LT603828-35	LT603933-38	LT604040-47
839	A	Septiembre-2006	LT603836-43	LT603939-45	LT604048-55
1221	B	Febrero-2009	LT603844-51	LT603946-58	LT604056-65
1343	B	Febrero-2010	LT603852-59	LT603959-66	LT604064-71

Nota. ENA: *European Nucleotide Archive*

3.1.1.4 Amplificación

Las amplificaciones de los tres loci se realizaron utilizando cebadores diseñados en regiones conservadas. En los casos en que éstos se mostraron ineficaces a la hora de producir secuencias de los ensamblajes esperados a tenor de los resultados del genotipado con *tpi*, se utilizaron los cebadores específicos para cada ensamblaje (Tabla 3.1).

Las PCR anidadas suelen implicar un mayor número de ciclos de PCR, lo que incrementa la posibilidad de que se produzcan errores de la polimerasa durante la amplificación. Para minimizar este efecto se utilizó una mezcla de enzimas de alta fidelidad (FastStart High Fidelity PCR System; Roche Diagnostics) con una tasa de error muy reducida respecto a las Taq polimerasas más comunes.

Las reacciones de PCR se prepararon como se describe en la Tabla 3.5 y las condiciones de amplificación fueron las mismas que las descritas en la Tabla 3.3 ajustando las temperaturas de anillamiento para cada pareja de cebadores como se describe en la Tabla 3.1.

Tabla 3.5: Condiciones de las reacciones de PCR para secuenciación.

	PCR externa	PCR interna
Tampón Fast Start High Fidelity 10x	1 µL	2 µL
DMSO	0.4 µL	0.8 µL
dNTPs	200 µM	200 µM
Cebadores	0.4 µM <i>Forward</i> 0.4 µM <i>Reverse</i>	0.4 µM <i>Forward</i> 0.4 µM <i>Reverse</i>
Fast Start High Fidelity ADN polimerasa	1 U	2 U
ADN	1-2 µL	1 µL de producto de la PCR externa *
H ₂ O	hasta 10 µL	hasta 20 µL
Volumen Total	10 µL	20 µL

Nota: * diluido 1:20 o sin diluir en función de la eficacia de la amplificación.

3.1.1.5 Comprobación de los ensamblajes con enzimas de restricción

Tras la amplificación de las muestras mixtas con cebadores específicos se comprobó que los fragmentos obtenidos pertenecían al ensamblaje esperado mediante la digestión de los productos de PCR con enzimas de restricción. Estas enzimas cortan la cadena nucleotídica en determinadas posiciones en función de la secuencia, produciendo fragmentos de distintos tamaños que se pueden visualizar mediante electroforesis en geles de agarosa. Los amplicones de *gdh* pertenecientes al ensamblaje A presentan tres dianas de restricción de AluI (posiciones 450, 741 y 750 respecto al codón de inicio) mientras que en el ensamblaje B, hay una cuarta diana en la posición 634 que genera dos fragmentos de menor tamaño. AluI también corta en distintas posiciones los amplicones de *calt* generando cuatro fragmentos en el ensamblaje A y dos en B. Por su parte la enzima de restricción BglI corta los amplicones de *bg* del ensamblaje A en la posición 431 pero no los del ensamblaje B.

3.1.1.6 Clonación

Los productos de amplificación se visualizaron mediante electroforesis en geles de agarosa (1.5% p/v, tampón Tris-Bórico-EDTA, TBE) teñidos con bromuro de etidio (5% v/v). Los amplicones de tamaño esperado se cortaron y se extrajeron de la matriz del gel utilizando el QIAquick Gel Extraction Kit (QIAGEN) para clonarlos, a continuación, con diversos kits comerciales (TOPO TA Cloning Kit for Sequencing, Invitrogen; TA Cloning Kit, Invitrogen; o CloneJET PCR Cloning Kit, Thermo Scientific). Una vez ligados los productos, los vectores fueron transformados en células competentes One Shot TOP10 *Escherichia coli* (Invitrogen). Todas las reacciones de clonación se escalaron a la mitad del volumen propuesto por el fabricante.

Las células transformadas se sembraron en un medio selectivo LB-Agar con ampicilina (100 µg/mL). Los vectores de los kits TOPO TA y CloneJET contienen un gen letal que mata las células en caso de que el plásmido no tenga inserto, mientras que el vector incluido en el kit TA utiliza el sistema lacZ para la selección de las colonias con inserto por lo cual las placas tuvieron que ser tratadas con 40 µL de X-Gal (40 mg/mL). La comprobación del tamaño del inserto en las colonias positivas se llevó a cabo mediante una PCR de cribado utilizando la Dream Taq Polymerase (Thermo Scientific) y los cebadores universales M13 o

pJET1.2 incluidos en el vector correspondiente. Las colonias se resuspendieron en 10 µL de agua, se inactivaron (10 minutos a 94°C en 10 µL de agua) y 1 µL se utilizó como molde de PCR. El programa utilizado fue el siguiente: desnaturalización inicial de 2' a 94°C; 35 ciclos de 30'' a 94°C, 30'' a 55°C y 1' a 72°C; seguido por una extensión final de 5' a 72°C.

3.1.1.7 Secuenciación

Se secuenciaron un mínimo de ocho colonias por ensamblaje y muestra. La probabilidad de no encontrar un alelo representado a una frecuencia p al examinar un número de secuencias n es $P = (1-p)^n$, de modo que al secuenciar un mínimo de 8 colonias la probabilidad de no encontrar un alelo a una frecuencia del 50% es $P = (1 - 0.5)^8 < 0.005$.

Para la amplificación de los clones se realizó una PCR con cebadores M13 F/R y FastStart High Fidelity Polymerase empleando el programa descrito en la Tabla 3.3 y utilizando como ADN molde 1 µL de la solución de colonias inactivadas.

Los productos de PCR se secuenciaron en sentido directo y reverso en la Unidad de Secuenciación y Análisis de fragmentos de la Universidade de Santiago de Compostela usando la química BigDye® Terminator v3.1 en el secuenciador capilar ABI 3730xl (Applied Biosystems).

Las secuencias se revisaron utilizando el software CodonCode Aligner (CodonCode Corporation, Dedham, MA, USA) para la correcta asignación de bases nucleotídicas, se alinearon con MUSCLE (Edgar, 2004) y se editaron manualmente con BioEdit (Hall, 1999). Las secuencias resultantes se depositaron en el European Nucleotide Archive (ENA), números de acceso: LT603736-LT604071 (Tabla 3.4). Dos secuencias que contenían codones de parada inesperados no pudieron ser depositadas en la base de datos.

3.1.2 Análisis genético-poblacional

3.1.2.1 Diversidad y divergencia

Los estudios de polimorfismo de ADN se realizaron con DnaSP v5.10.01 (Librado y Rozas, 2009) y MEGA versión 5 (Tamura *et al.*, 2011). Las regiones no solapantes de los alineamientos, así como aquellas correspondientes a las secuencias de los cebadores, se

excluyeron del análisis. La diversidad haplotípica (H_d) se calculó como la probabilidad inversa de que dos secuencias al azar en la muestra fuesen idénticas (Nei, 1987). Para la determinación de la diversidad nucleotídica se utilizaron los estadísticos π (Nei, 1987) y θ_w (Watterson, 1975) para posiciones sinónimas (π_s y θ_s) y no sinónimas (π_A y θ_A), aplicando la corrección de Jukes-Cantor (JC) (Jukes y Cantor, 1969). π estima el número promedio de diferencias que presentan dos secuencias extraídas al azar en una población. En esta tesis su valor se expresa como porcentaje, es decir, por cada 100 pares de bases (pb). Por su parte, θ_w estima el número total de posiciones segregantes en la muestra sin tener en cuenta su frecuencia; también se expresa como porcentaje.

Además, para verificar si las mutaciones detectadas se ajustaban a lo esperado bajo la hipótesis de neutralidad (Kimura, 1983) se utilizó la prueba D de Tajima, que compara las estimas de diversidad π y θ_w (Tajima, 1989). En condiciones de equilibrio mutación-deriva ambos estadísticos son buenas estimas de la diversidad genética y presentan valores similares, siendo el valor de D próximo a cero. Valores negativos de D indican un exceso de variantes a baja frecuencia ($\theta_w > \pi$) mientras que valores positivos indican un exceso de variantes a frecuencias intermedias ($\pi > \theta_w$).

La divergencia nucleotídica (K = número de sustituciones sinónimas por posición) entre los ensamblajes A y B se estimó de forma separada en lugares sinónimos (K_s) y no sinónimos (K_A), utilizando el método de Nei-Gojorobi para calcular el número de posiciones sinónimas y no sinónimas y aplicando la corrección de JC (Nei y Gojobori, 1986).

Los análisis de diversidad se llevaron a cabo tanto en el conjunto de aislados de cada ensamblaje como dentro de cada aislado. Las regiones no solapantes se excluyeron de los alineamientos para que la comparación de los datos entre aislados y ensamblajes no se viera afectada por la diferencia de longitud de las secuencias en función de los cebadores empleados para cada amplificación.

3.1.2.2 Estructura poblacional

Para analizar la estructuración genética de la población se realizó un análisis de varianza molecular (AMOVA) utilizando el paquete informático Arlequin 3.5 (Excoffier y Lischer, 2010). Este método combina información de la varianza de las frecuencias haplotípicas y la

distancia genética entre haplotipos para cuantificar los distintos componentes de la covarianza correspondientes con los distintos niveles de estructura genética en la población (en este caso: dentro de cada hospedador, entre hospedadores y entre grupos de hospedadores). El nivel de significación se calculó utilizando permutaciones paramétricas.

3.1.2.3 Recombinación

La tasa de recombinación a escala poblacional (ρ) se estimó mediante el método de verosimilitud compuesta de Hudson (2001) siguiendo el modelo de lugares finitos implementado en el paquete informático LDhat v2.2 (McVean *et al.*, 2002). *Giardia* es un organismo tetraploide, por lo que ρ es igual al doble de $4N_e$ multiplicado por la tasa de recombinación ($\rho = 8N_e r$), siendo N_e el tamaño efectivo de población.

LDhat también realiza la prueba de cuatro gametos ($G4$), que detecta eventos de recombinación en caso de que dos posiciones segregantes bialélicas muestren las cuatro combinaciones haplotípicas posibles (Hudson y Kaplan, 1985), así como otras dos estimas que calculan la correlación entre las medidas de desequilibrio de ligamiento $|D'|$ (Lewontin, 1964) y r^2 (Hill y Robertson, 1968) y la distancia física entre pares de lugares segregantes ($\text{corr}(r^2, d)$ y $\text{corr}(|D'|, d)$, respectivamente). Ambas se basan en la hipótesis nula de que, en ausencia de recombinación, el desequilibrio de ligamiento es constante en todo el genoma e independiente de la distancia entre dos loci, mientras que la probabilidad de que se produzcan eventos de recombinación que erosionen el desequilibrio de ligamiento entre dos posiciones segregantes es mayor cuanto más alejadas estén entre ellas.

Para la determinación del nivel de significación de los valores obtenidos, LDhat calcula la distribución nula de cuatro estadísticos (verosimilitud compuesta máxima, sumatorio de las diferencias entre pares de lugares segregantes detectados por $G4$, $\text{corr}(r^2, d)$ y $\text{corr}(|D'|, d)$) bajo permutación aleatoria (1000 permutaciones) de la posición física de los lugares segregantes y la compara con los valores observados en la muestra sin permutar (McVean *et al.*, 2002).

El programa RecMin (Myers y Griffiths, 2003) se utilizó para calcular el número mínimo de eventos de recombinación. Este se estima en función de la diferencia entre el número de haplotipos en la muestra y el número de posiciones segregantes (R_h), asumiendo un modelo

de posiciones infinitas donde cada posición segregante ha sufrido un único evento de mutación desde el ancestro común más reciente del conjunto de individuos analizado.

3.1.2.4 Análisis filogenético

El análisis filogenético se llevó a cabo utilizando MEGA versión 5 (Tamura *et al.*, 2011). Las distancias evolutivas entre alelos se calcularon siguiendo el modelo de JC y los árboles se construyeron utilizando el método Neighbor-Joining (Saitou y Nei, 1987). La fiabilidad de la topología resultante se evaluó mediante la técnica de múltiples réplicas o *bootstrap* ($N = 500$) (Felsenstein, 2009).

En los alineamientos empleados para la construcción de las filogenias se incluyeron los distintos haplotipos detectados para cada locus, las secuencias de los genomas de referencia (aislado GS para B, aislado WB para AI y aislado DH para AII) y, en el caso de *gdh* y *bg*, secuencias publicadas en estudios anteriores (Lalle *et al.*, 2005; Wielinga y Thompson, 2007). Además se incluyeron secuencias de grupos externos que, para *gdh*, *bg* y *calt*, fueron *Giardia ardeae* (AF069060), *Giardia muris* (EF455599) y *Spironucleus salmonicida* (KI546059), respectivamente.

3.1.2.5 Redes de haplotipos

A diferencia de lo que ocurre en los árboles filogenéticos, que analizan el patrón de evolución dicotómica más probable en función de las distancias genéticas entre secuencias (basándose en un modelo direccional por acumulación de mutaciones), las redes reflejan de una forma aproximada como se relacionan todos los haplotipos entre sí.

La redes de haplotipos se construyeron siguiendo el método *Median-Joining* (Bandelt *et al.*, 1999) con el programa NETWORK v5.0.0.0 (disponible en www.fluxus-engineering.com), utilizando los parámetros por defecto. Se incluyeron las secuencias descritas en este trabajo junto con otras secuencias solapantes de aislados humanos, procedentes de GenBank, cuyo origen geográfico pudo ser verificado. En total, 137 secuencias de *gdh*-A (posiciones 457-634 con respecto al codón de inicio), 264 de *gdh*-B (posiciones: 459-641), 105 de *bg*-A (posiciones: 217-539) y 174 de *bg*-B (posiciones: 156-539).

3.2 EXPERIMENTO II: GENOTIPADO MULTILOCUS DE QUISTES INDIVIDUALES

3.2.1 Aislamiento de quistes individuales

3.2.1.1 Selección de muestras

Para el análisis de quistes individuales se utilizaron cinco muestras de pacientes del Área Sanitaria de Santiago de Compostela recogidas entre noviembre de 2014 y mayo de 2015. En este caso fue necesario el uso de muestras frescas con una concentración elevada de quistes viables para que los procesos posteriores se pudiesen llevar a cabo con éxito. Dichos quistes mantienen su forma redondeada y una elevada refringencia al ser observados en el microscopio de contraste de fases, en comparación con quistes rotos o muertos.

El análisis se centró exclusivamente en aislados del ensamblaje B, ya que los aislados de este ensamblaje muestran una mayor heterogeneidad genética. Esto los convierte en un modelo más adecuado para estudiar la estructura genética de las infecciones de *Giardia* en humanos, ya que a mayores niveles de diversidad mayor poder de resolución de la técnica para detectar diferencias dentro o entre individuos. Este razonamiento es válido igualmente para el estudio de la recombinación, ya que cuanto más variabilidad exista dentro de cada locus más sencillo resultará detectarla.

3.2.1.2 Purificación y concentración de los quistes

Antes de la micromanipulación se procesaron las muestras de heces para obtener una suspensión purificada y concentrada de quistes utilizando un gradiente de sacarosa. Brevemente, el protocolo utilizado fue el siguiente:

1. Entre 2-5 g de muestra fecal se resuspendieron en ~5 mL de solución salina (NaCl 0,5%) y se filtraron a través de una gasa húmeda. Esta suspensión pasó a un tubo de 10 mL y se centrifugó a 250 g durante 4 min en una centrífuga de sobremesa (Jouan C411). Se desechó el sobrenadante.

2. El precipitado se lavó en 10 mL de una solución de Tween 80 al 0,5% (centrifugación: 250 g, 4 min). Este proceso se repitió 3-4 veces hasta que el sobrenadante quedó limpio.
3. El precipitado se resuspendió en 5 mL de agua destilada y se depositó sobre 5 mL de una solución de sacarosa 0,85 M en un nuevo tubo de 10 mL, con especial cuidado de no producir turbulencias en la interfase entre ambas soluciones. Centrifugación a 500 g durante 20 min.
4. Los quistes se recuperaron de la interfase agua-sacarosa con ayuda de una pipeta Pasteur, se lavaron en tampón fosfato salino (PBS, BioMérieux) para eliminar los restos de sacarosa (centrifugación: 250g, 4 min) y finalmente se resuspendieron en 10 mL de PBS.

3.2.1.3 Micromanipulación

El aislamiento de los quistes se llevó a cabo con un microinyector IM-9B unido al micromanipulador MMO-202ND (Narishige). Las agujas capilares se prepararon en el laboratorio utilizando una forja Narishige.

El protocolo seguido fue el siguiente:

1. Antes de iniciar la micromanipulación se dejó que sedimentaran los quistes, se retiró el sobrenadante, y se resuspendieron en 10 mL de PBS.
2. En una placa petri se colocó un cubreobjetos limpio sobre el que se dispuso una gota de 400 μ L de la suspensión de quistes. Se añadió PBS hasta cubrir la superficie de la placa.
3. La preparación se observó a 100 y 200 aumentos en un microscopio invertido de contraste de fases (Nikon, Diaphot 200) y con ayuda del micromanipulador se procedió a extraer quistes individualmente.
4. Cada quiste se lavó dos veces por repipeteo en sendas gotas de 400 μ L de PBS dispuestas sobre dos portaobjetos limpios en una nueva placa petri.
5. Finalmente, se colocaron dos gotas de 1 μ L de agua sobre dos pequeños fragmentos recortados de un cubreobjetos; en una se liberó el volumen extra de líquido de suspensión y en la otra se depositó el quiste. Ambos fragmentos de vidrio se transfirieron a sendos tubos de PCR con ayuda de unas pinzas.

Además de los quistes aislados, también se preparó un control positivo de PCR consistente en 1 μ L de la suspensión de quistes para cada muestra (>100 quistes) que igualmente fue depositado en un fragmento de vidrio y transferido a un tubo de PCR.

3.2.2 Amplificación directa multiplex

3.2.2.1 Selección de marcadores y diseño de cebadores

La selección de marcadores se realizó en función de la disponibilidad de secuencias del ensamblaje B en las bases de datos. Se incluyeron siete genes nucleares, de copia única, conservados, no sometidos a selección natural y que se encuentra repartidos en los cinco cromosomas del genoma de *G. duodenalis* (aislado WB): los tres loci más comúnmente utilizados en el genotipado de *Giardia*: *bg*, *tpi* y *gdh*; otros tres genes menos frecuentemente estudiados, *rpL7a* (GL50803_17244), *mlh1* (GL50803_16149) y 2Fe-2S *fd* (GL50803_27266); y *calt* del cual teníamos datos propios del primer experimento. Los cebadores utilizados para la amplificación de cada uno de estos loci y los tamaños de esperados de los amplicones se indican en la Tabla 3.6.

Tabla 3.6: Cebadores PCR multiplex.

Cebador	Secuencia (5'→ 3')	Longitud del amplicón
bgF	CGCGTGTGAAGATGATCAAG	306 pb
bgR	CTGTCTCGAGGTCGTTTCAGG	
tpiF	TCCCTTCATCGGTGGTAACT	332 pb
tpiR	GCTTCGCCTTCTTAGCACTC	
RPL7F	GTCGAAGGAAGAGCACAAGG	356 pb
RPL7R	TTGGTGCGGACAATAGCATA	
mlhF	ATGTCACCGCTTTGCAACTA	360 pb
mlhR	GCAGCAGAAATGGACATCAA	
caltF	AGCATATCAACGGGGAAGC	378 pb
caltR	ATTCTCCGACAATTCCTTGG	
fdF	GCAGGGCGTTTCATAACATT	379 pb
fdR	GCTGTCCGCACTTCTTCTTT	
gdhF	CTAAGGGCAAGTCGGACAAC	391 pb
gdhR	TTGGAGTCCGAGAAGGTGAG	

Para el diseño de cebadores se utilizaron las mismas herramientas que en el primer experimento. Los resultados obtenidos con Primer-Blast se analizaron con Autodimer para examinar la posible formación de dímeros entre cebadores. El tamaño máximo de los amplicones se ajustó a las necesidades del método de secuenciación que ofrece resultados óptimos con secuencias de hasta 400 nucleótidos. En la medida de lo posible, los cebadores se diseñaron para que los amplicones fuesen de distintos tamaños. Sin embargo, para no perder demasiada información, el tamaño mínimo de los amplicones se estableció en 300 pb, lo que conllevó que los amplicones de *fd* y *calt*, así como los de *rpL7a* y *mlh1* tuviesen tamaños similares.

3.2.2.2 PCR multiplex

El sistema de PCR utilizado –Phusion Direct Human Specimen Kit (Thermo Scientific)– permite amplificar muestras biológicas sin necesidad de extracción y purificación previa del ADN. Esta propiedad evita cualquier tipo de manipulación de los quistes desde que son aislados hasta la PCR, lo que minimiza tanto el riesgo de contaminación como el riesgo de perder alelos, en ambos casos factores importantes dado el mínimo material genético de partida (cada quiste incluye 16 copias del genoma).

La reacción de PCR se preparó en un volumen total de 10 μ L, como se indica en la Tabla 3.7. La mezcla se añadió directamente sobre las secciones de vidrio donde se habían depositado los quistes (y sus controles) dentro de cada tubo de PCR. Las condiciones de amplificación se muestran en la Tabla 3.8.

Tabla 3.7: Condiciones de la reacción de PCR multiplex

	Volumen
Tampón Phusion Human Specimen 2X	5 μ L
Cebadores	1 μ L*
ADN polimerasa Phusion Hot Start II High-Fidelity	0,2 μ L
Quiste	1 μ L
H ₂ O	2,8 μ L

Nota: *concentración final de los cebadores: 500 μ M para *bg*, *fd*, *calt*, *mlh1* y *tpi* y 400 μ M para *RPL7a* y *gdh*

Tabla 3.8: Condiciones de amplificación de la PCR multiplex.

	Tiempo	T ^a	Ciclos
Desnaturalización inicial	5'	98°C	1
Desnaturalización	5''	98°C	35
Anillamiento	15''	65°C	
Extensión	20'	72°C	
Desnaturalización	5''	98°C	10
Anillamiento	15''	63°C	
Extensión	20'	72°C	
Extensión final	1'	72°C	1

Los productos de PCR se visualizaron mediante electroforesis en geles de agarosa al 2.5% p/v (TBE) teñidos con bromuro de etidio al 5% v/v, utilizando el volumen total de la PCR en el caso de los controles negativos y 2 µL en el caso de los quistes y los controles positivos. En la mayor parte de las amplificaciones de quistes individuales la banda de *tpi* fue muy tenue y se llevó a cabo una reamplificación específica de este locus (15 ciclos y temperatura de anillamiento de 65 °C) utilizando como molde 1 µL de producto de la PCR multiplex. La concentración de los amplicones se estimó de modo aproximado a partir de la intensidad de la tinción de las bandas. La reamplificación de *tpi* se añadió a la multiplex a concentración equimolar.

Finalmente, se seleccionaron ocho quistes y un control positivo por muestra para su secuenciación. De esta forma, la probabilidad de detectar un genotipo que se encuentra a una frecuencia del 25% o superior en la muestra es > 0.9 .

3.2.3 Secuenciación masiva en paralelo

Las plataformas de secuenciación masiva en paralelo se basan en la secuenciación independiente de un elevado número de moléculas de ADN de forma simultánea. Para poder aislar y amplificar la señal que produce cada una de estas moléculas, utilizan distintas estrategias que permiten su amplificación clonal sobre un soporte sólido. De modo que se generan clones aislados físicamente que son secuenciados por separado.

En este estudio se utilizó la tecnología Ion Torrent (Rothberg *et al.*, 2011), que se basa en la detección de variaciones de pH durante la polimerización del ADN. La amplificación clonal de las moléculas de ADN se produce sobre unas partículas esféricas (ISP, Ion SphereTM particles) mediante PCR en emulsión, para lo cual es necesario que cada molécula lleve en sus extremos una pareja de adaptadores universales que sirven como cebadores.

El flujo de trabajo desde el aislamiento de los quistes hasta la secuenciación consta de cuatro etapas: i) captura de las regiones de interés (ver apartado anterior); ii) preparación de librerías de amplicones; iii) preparación del molde de secuenciación y iv) secuenciación (Figura 3.1).

3.2.3.1 Construcción de librerías de amplicones

Una librería de amplicones es un conjunto de moléculas de ADN obtenidas de una misma amplificación de PCR (en este caso un mismo quiste o control de aislado) que presentan en sus extremos los adaptadores universales A y P1 (Figura 3.1). Estos adaptadores se unen mediante ligación y son necesarios para la amplificación clonal de las moléculas de ADN sobre las ISP y en la reacción de secuenciación. Para poder incluir los amplicones de todos los quistes en una misma carrera de secuenciación, cada librería de amplicones se marcó con un código de barras distinto (CB, Figura 3.1). Estos códigos son unas secuencias específicas de entre 10 y 12 nucleótidos de longitud, situadas a continuación del adaptador A, que permiten identificar la librería de origen de cada molécula de ADN en la secuenciación (Figura 3.1).

En la preparación de la librerías se utilizaron el Ion Plus Fragment Library Kit (Life Technologies) y los adaptadores 1-45 del Ion XpressTM Barcode Adapters 1-96 Kit (Life Technologies), siguiendo las instrucciones del fabricante (revisión A.0 del protocolo “Prepare Amplicon Libraries without Fragmentation Using the Ion Plus Fragment Library Kit”). La construcción de cada librería se inició con un máximo de 100 ng de productos de PCR diluidos en un volumen de 50 µL. La concentración de ADN se midió utilizando el Agilent D1000 ScreenTape System en una Agilent 2200 TapeStation. Se purificaron los amplicones, se repararon los extremos de los fragmentos y se ligaron los adaptadores. Dos purificaciones (en lugar de una como establece el protocolo) permitieron eliminar los adaptadores que no se ligaron a ninguna cadena de ADN. Por último se cuantificaron las librerías mediante qPCR y se mezclaron a concentración equimolar (Figura 3.1).

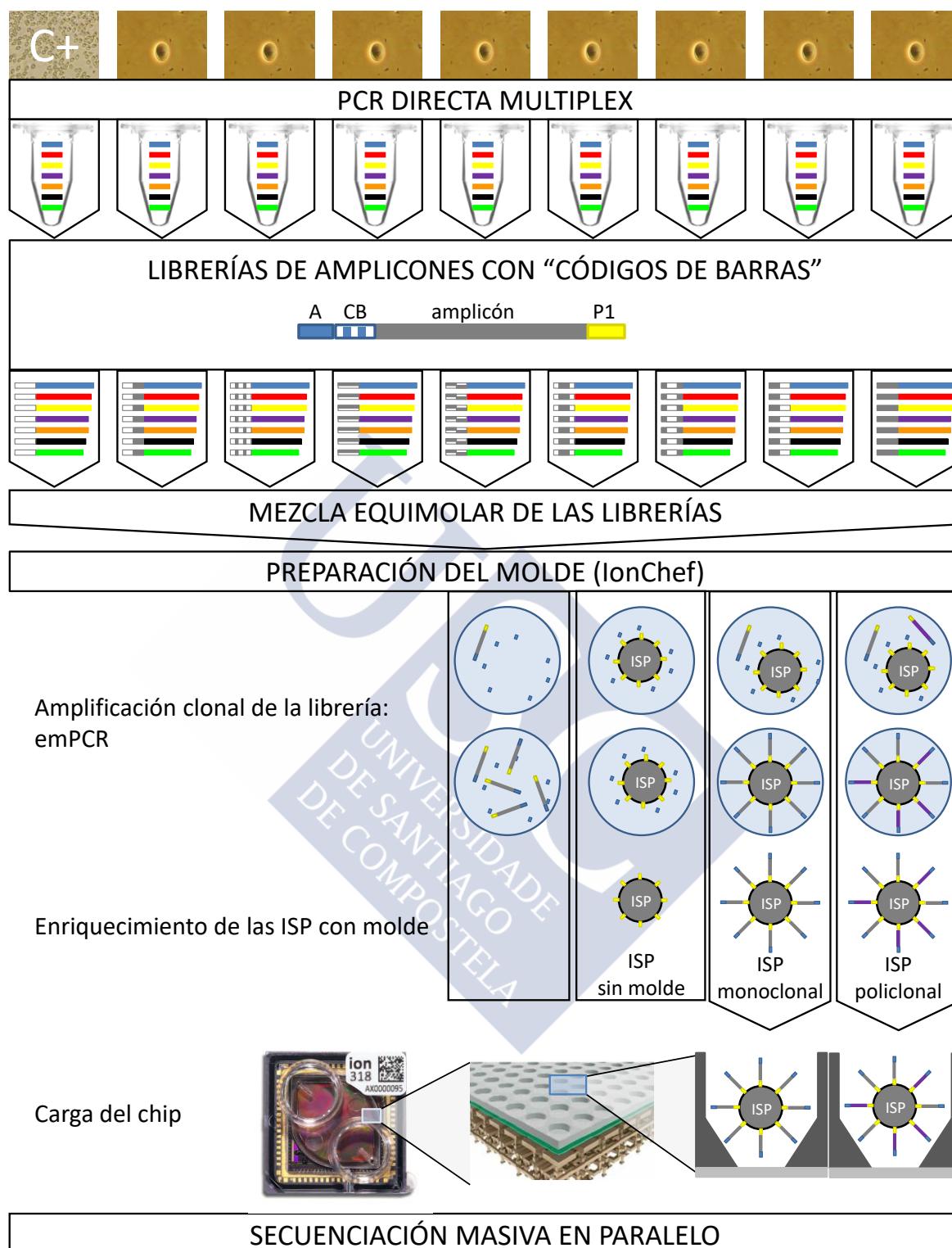


Figura 3.1: Representación esquemática del proceso desde el aislamiento de los quistes hasta la MPS. Se construyeron librerías de amplicones de cada quiste y control (ligación de adaptadores A y P1) marcadas con distintos códigos de barras (CB) y se mezclaron a concentración equimolar. La librería de secuenciación resultante se amplificó clonalmente sobre las ISP (emPCR). Las ISP sin molde se descartaron antes de cargar el chip de secuenciación. Las ISP policlonales se secuenciaron y se descartaron en el análisis de datos.

3.2.3.2 Preparación del molde de secuenciación

La preparación del ADN molde para la secuenciación consta de tres etapas: i) la amplificación clonal de las moléculas de ADN sobre las ISP, ii) el enriquecimiento de la fracción de ISP sobre las que se produjo amplificación y iii) la carga del chip. Estos tres procesos se llevaron a cabo de forma automatizada en el Ion ChefTM System (ThermoFisher), utilizando el Ion PGMTM Hi-Q Chef Kit (ThermoFisher), adaptado para la secuenciación de fragmentos de 400 pb.

La emPCR permite aislar las reacciones de amplificación en microgotas acuosas en una matriz de aceite. Idealmente cada microgota contiene una ISP cubierta de adaptadores P1, una molécula de ADN con adaptadores A y P1 en sus extremos y adaptadores libres que sirven como cebadores. La concentración de la librería de secuenciación se optimizó para que cada ISP coincidiese en una microgota con una o ninguna molécula de ADN, siguiendo una distribución de Poisson (Figura 3.1). Sin embargo, esta distribución predice que una fracción menor de las ISPs entrará en contacto con más de una molécula de ADN.

Cuando una ISP no entra en contacto con ninguna molécula de ADN, o lo hace con una molécula que no tiene los adaptadores ligados correctamente (adaptadores A o P1 en ambos extremos), se genera una ISP sin molde. Si entra en contacto con una única molécula se genera una ISP monoclonal y si lo hace con dos moléculas distintas se genera una ISP policlonal (Figura 3.1). Las ISPs sin molde se descartan antes de la carga del chip de secuenciación en la fase de enriquecimiento de ISP, pero las ISPs policlonales no se detectan hasta la secuenciación y las lecturas que generan son descartadas posteriormente.

Para la secuenciación se utilizó un chip Ion-318v2 (ThermoFisher) con ~11.2 millones de micropocillos en su superficie. En cada uno de estos pocillos se puede depositar una única ISP, lo que aísla la señal que produce cada conjunto de moléculas clonales. Cada pocillo está provisto de un detector de variación de pH individual, lo que permite la secuenciación independiente en paralelo de millones de ISPs (Figura 3.1).

3.2.3.3 Secuenciación

La secuenciación de los productos de la emPCR se realizó en un Ion PGMTM (Ion Torrent Personal Genome Machine, ThermoFisher; Rothberg *et al.*, 2011), utilizando una química Ion PGMTM Hi-Q Sequencing Kit para fragmentos de 400 pb (ThermoFisher). Esta tecnología de secuenciación por síntesis se basa en la detección de variaciones en el pH durante la polimerización del ADN: cuando un nuevo nucleótido se añade a la cadena de ADN se forma un enlace fosfodiéster y se liberan un pirofosfato y un H⁺, lo que provoca un descenso del pH en los pocillos donde tiene lugar la reacción de polimerización. La variación de pH es detectada por un sensor situado en la base del pocillo, transformada en una variación de voltaje y registrada (Figura 3.2).

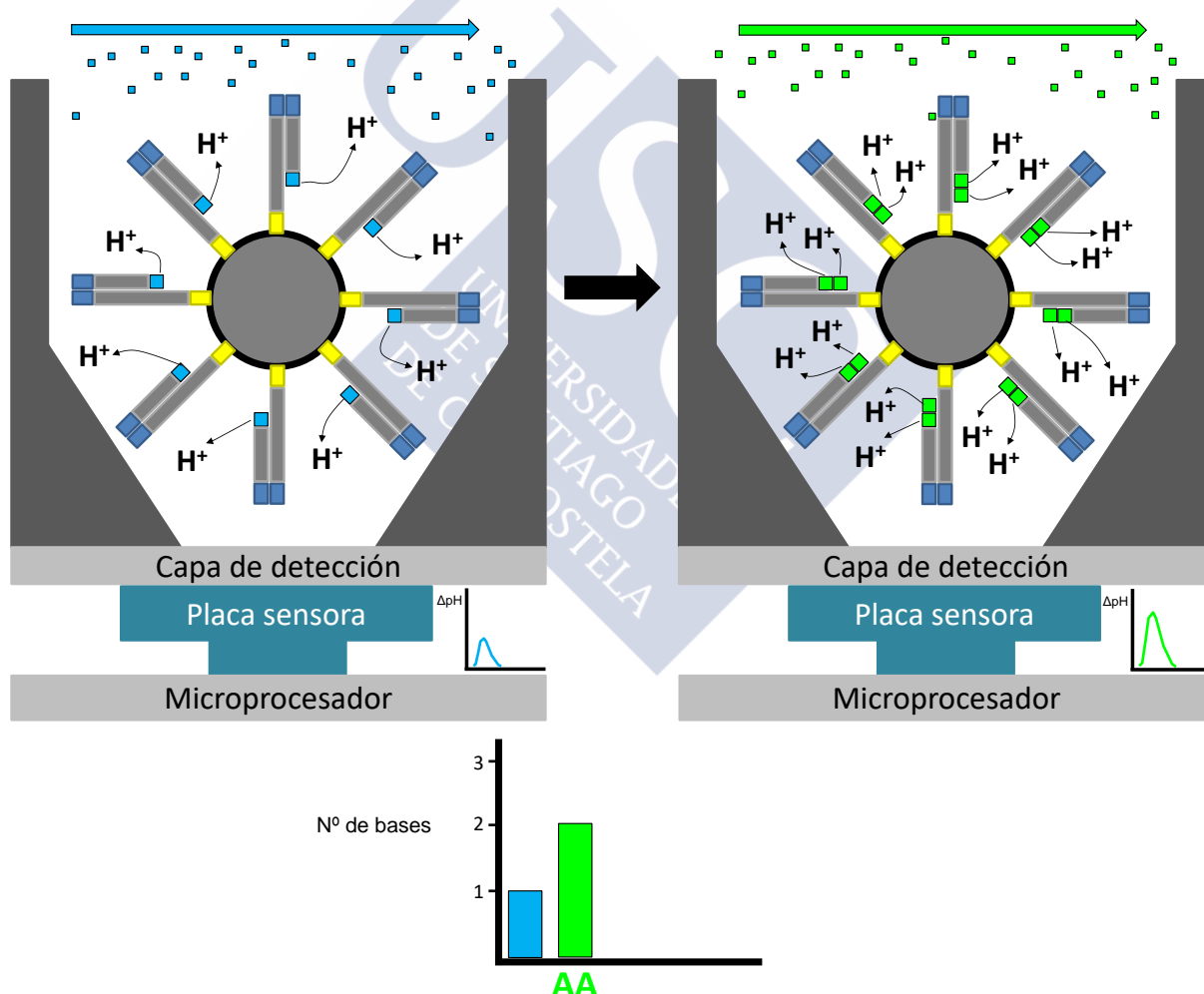


Figura 3.2: Esquema de la reacción de secuenciación en un micropocillo del chip semiconductor. El adaptador A (azul oscuro) sirve como cebador para iniciar la polimerización de la cadena de síntesis. Las flechas azul y verde indican flujos consecutivos de soluciones de dCTP y dATP (con un lavado entre ambos).

Durante la secuenciación, los cuatro dNTPs se añaden al chip en ciclos alternos, inundando los pocillos. En aquellos en los que corresponde por complementariedad de bases con la cadena molde, se produce la adición nucleotídica. En cada ciclo, el chip semiconductor detecta en qué pocillos se produce polimerización y, en función de la intensidad de la señal, estima cuántos nucleótidos se han añadido a la cadena de síntesis. De esta forma se establece la secuencia nucleotídica del clon de moléculas de ADN capturadas en cada ISP (Figura 3.2).

Los datos de secuenciación se procesaron con Torrent SuiteTM 4.4.3, un programa informático que compiló las lecturas, las filtró –descartando ISP policlonales, dímeros de adaptadores, lecturas de baja calidad y fragmentos control–, las asignó a cada quiste en función del código de barras y las alineó contra el genoma de referencia del ensamblaje B (aislado GS; GenBank *accession number*: GCA_000498735). Los alineamientos resultantes se inspeccionaron utilizando la herramienta de visualización de genomas IGV (Integrative Genomics Viewer; Robinson *et al.*, 2011). Los cálculos del número de lecturas por locus y por quiste/muestra y la tasa de inserciones y deleciones (*indels*) se realizaron con SAMTools (Li *et al.*, 2009) y el análisis de la cobertura se realizó con BEDTools (Quinlan y Hall, 2010). La profundidad de cobertura promedio se estimó como la suma de las lecturas con *Phred-like Quality score* > 30 (precisión de lectura del 99.9%) para cada posición, dividida por la longitud del amplicón (sin tener en cuenta los cebadores). La tasa de *indels* se calculó como el número de eventos en un alineamiento, dividido por la longitud del mismo.

3.2.4 Análisis de secuencias

3.2.4.1 Identificación de alelos

Las lecturas generadas mediante MPS son clonales, es decir, se obtienen a partir de una única molécula de ADN, lo que permite asociar cada una de ellas con un haplotipo: una combinación de variantes en la misma molécula de ADN que se heredan de forma conjunta. Sin embargo, las plataformas de NGS en general, e Ion PGMTM en particular, presentan una elevada tasa de error que se ve compensada por la enorme cantidad de datos que generan (Bragg *et al.*, 2013). En el caso de la secuenciación basada en semiconductores el mayor problema es la secuenciación de homopolímeros, durante la cual se generan inserciones o deleciones espurias de un número reducido de nucleótidos (*indels*) a una tasa muy elevada.

Además, la mayor sensibilidad de esta tecnología respecto a métodos clásicos de secuenciación, provoca que se detecten artefactos que se producen a baja frecuencia durante la PCR (recombinantes de PCR y mutaciones espurias por errores de la polimerasa) y la hace más susceptible a bajos niveles de contaminación entre muestras. Esto provoca que muchas de las lecturas contengan algún error de secuenciación y por tanto no se correspondan con un haplotipo real, sino con un artefacto derivado de éstos.

Debido a estas particularidades, en los apartados referidos al experimento de MPS se utiliza “haplotipo” para referirse a cualquier variante haplotípica detectada durante la secuenciación, mientras que el término “alelo” se emplea para referirse a las variantes haplotípicas reales presentes en cada locus.

Para identificar los alelos presentes en cada quiste (genotipado de los quistes) se emplearon tres métodos distintos: el alineamiento y filtrado de secuencias únicas (USEQs), el programa informático AMPLIAN (Zagordi *et al.*, 2011) y la herramienta *online* AMPLISAS (Sebastian *et al.*, 2016). Aunque cada uno de estos métodos está pensado para dar un resultado definitivo, la comparación de los datos de los tres ofreció un resultado más completo y fiable, ya que unos cubrieron las debilidades de los otros. Por tanto, se consideraron alelos todos los haplotipos que pasaron los filtros de calidad y frecuencia que se exponen a continuación en al menos uno de los tres métodos.

3.2.4.1.1 Alineamiento de secuencias únicas (método USEQ)

El primer método es un simple protocolo de filtrado de secuencias únicas (USEQs), que se basa en el supuesto de que los artefactos de PCR y los errores de secuenciación son menos comunes que los alelos reales. Para obtener el alineamiento de USEQs se utilizó la siguiente metodología:

1. Las lecturas obtenidas para cada quiste/muestra se separaron por locus utilizando SAMTools.
2. Las lecturas con calidad de mapeo MAPQ>60 de un tamaño superior al 95% del tamaño de amplicón esperado se extrajeron en archivos de líneas de texto.

3. Las lecturas idénticas se colapsaron en una única secuencia contando en cada caso el número de lecturas: USEQ.

```
>for file in *.txt; do sort -k 2 $file | uniq -c > $file.useq.txt
```

4. Las USEQs con una única lectura se descartaron, ya que se corresponden con errores o combinaciones de errores tan poco comunes que ni siquiera se repitieron.

En este paso, los archivos se convirtieron a formato FASTA, se alinearon utilizando MUSCLE (*gap penalty* = -200) y se inspeccionaron en BioEdit.

5. Las USEQs de un tamaño distinto al esperado para cada amplicón se descartaron por contener *indels*.
6. A continuación, se descartaron las USEQs con un número de lecturas inferior al 7,5% de las lecturas acumuladas que cumplen los criterios anteriores.
7. Estos archivos de secuencias únicas se convirtieron a formato FASTA y se alinearon usando MUSCLE (*gap penalty* = -200) para ser inspeccionados visualmente con BioEdit. Se descartaron las secuencias que contenían *indels* que se contrarrestaban de modo que su longitud era igual a la del amplicón y todas aquellas secuencias que se podían explicar como recombinantes de dos secuencias con mayor número de lecturas.
8. Por último, se revisaron las mutaciones en IGV para comprobar si estaban sobrerrepresentadas en un sentido de secuenciación respecto al otro (*strand bias*), en cuyo caso se consideraron artefactos introducidos por el secuenciador.

Los criterios para descartar los supuestos artefactos se establecieron en función de las propiedades esperadas para los alelos de quistes individuales. En primer lugar, se espera que los alelos se encuentren igualmente representados en ambos sentidos de las lecturas (*forward* y *reverse*), por lo que aquellas mutaciones que aparecieron únicamente en un sentido de la secuenciación, o mostraron un fuerte sesgo a favor de alguno de ellos, se consideraron artefactos y se descartaron. En segundo lugar, no se espera que las regiones analizadas presenten *indels* que alteren la pauta de lectura del gen. En el caso de *rpL7a*, que contiene un intrón, los *indels* más comunes se revisaron manualmente y todos ellos presentaban *strand bias*. En tercer lugar, las USEQs recombinantes aparecieron siempre que se detectaron en un

mismo quiste dos o más secuencias con más de una posición variable, y en ningún caso a mayor frecuencia que los alelos supuestamente parentales. Además, se detectaron eventos de recombinación entre cada par de variantes consecutivas, llegando a observarse un número de quimeras superior al número máximo de alelos esperados en un organismo tetraploide (cuatro). Por último, el filtro de frecuencia se aplicó en función de la frecuencia alélica mínima a la que se espera encontrar un alelo en un quiste: 25%. Los haplotipos que pasaron los filtros de calidad pero no alcanzaron el 30% de esta frecuencia mínima (un 7,5% total) se descartaron.

3.2.4.1.2 *AMPLIAN*

En segundo lugar se utilizó *AMPLIAN*, un programa diseñado para el análisis de amplicones, incluido en el paquete de software *ShoRAH* (Short Reads Assembly into Haplotypes; Zagordi *et al.*, 2011). Este programa genera un alineamiento múltiple de secuencias en la región del genoma correspondiente al amplicón y mediante un algoritmo de agrupamiento probabilístico corrige los errores, infiere los haplotipos y sus frecuencias y estima siguiendo un modelo Bayesiano la calidad de la reconstrucción de dichos haplotipos.

Los haplotipos con una probabilidad posterior inferior a 0,95 se descartaron por recomendación del autor del programa (comunicación personal). Los restantes se filtraron siguiendo los mismos criterios que en el método anterior: primero se eliminaron aquellos que presentaban *indels*, luego se filtró por frecuencia (7,5%) y finalmente se eliminaron las posibles quimeras y las mutaciones sesgadas. Este proceso se lleva a cabo por separado para cada locus y cada quiste.

3.2.4.1.3 *AMPLISAS*

AMPLISAS es una herramienta *online* de genotipado de amplicones a partir de datos de secuenciación masiva (<http://evobiolab.biol.amu.edu.pl/amplisat/index.php>; Sebastian *et al.*, 2016). Este programa realiza el genotipado de todos los loci y todas las muestras incluidas en la carrera de secuenciación en un único análisis. Se basa en el mismo supuesto que el método de *USEQS*: los alelos reales se encuentran a mayor frecuencia y las demás secuencias se corresponden con errores puntuales respecto a dichos alelos.

En primer lugar, las lecturas se asignan a la muestra correspondiente en función de su código de barras y se separan por locus. A continuación, se identifican los alelos más

comunes y los posibles artefactos de acuerdo a una serie de parámetros establecidos por el usuario en función de los resultados esperados y las características de la plataforma de secuenciación. Después, las secuencias se agrupan utilizando criterios de similitud y siguiendo otros parámetros como la frecuencia relativa de una secuencia respecto a la secuencia principal de un grupo (cuando una secuencia se encuentra a una frecuencia por encima de un cierto umbral dentro de un grupo se asigna como secuencia principal de un nuevo grupo). Finalmente, las secuencias principales de cada grupo se consideran alelos reales y se les asigna la frecuencia del conjunto. Los parámetros utilizados en este caso fueron: *minimum amplicon depth* = 50, *maximum allele number* = 4, *substitution threshold* = 0.1, *indel threshold* = 3, *cluster exact length* = Yes, *minimum dominant sequence frequency threshold* = 10%, *minimum amplicon sequence frequency* = 1%).

3.2.4.1.4 Confirmación por Sanger

Para tener un control independiente de los resultados del genotipado MPS, se secuenciaron por Sanger 16 clones por locus del aislado M (el que mostró mayores niveles de heterozigosidad en el análisis MPS). Cada locus se amplificó individualmente utilizando 1 µL de la suspensión de quistes bajo las mismas condiciones de PCR que las utilizadas para la amplificación multiplex. Los productos de PCR se clonaron siguiendo el procedimiento descrito en el Experimento I (apartado 3.1.1.6), los plásmidos se aislaron y purificaron utilizando el GeneJET plasmid miniprep Kit (Thermo Scientific) y se enviaron a GATC-Biotech (Constance, Alemania) para ser secuenciados.

El número de clones secuenciados se determinó en función de los resultados observados en el genotipado MPS. Los quistes de cada aislado presentaron los mismos genotipos, por lo que se consideró que los aislados representaron un conjunto de quistes clonales. En este caso se espera que los aislados presenten un máximo de cuatro alelos a una frecuencia mínima del 25%, al igual que ocurre en los quistes. Por tanto, siguiendo el principio de inclusión-exclusión, la probabilidad de no detectar los cuatro alelos en un locus tetralélico secuenciando n clones se obtiene de la ecuación: $P = \binom{4}{1}(3/4)^n - \binom{4}{2}(2/4)^n + \binom{4}{3}(1/4)^n$. Con $n = 16$ clones la probabilidad es $< 0,05$ y aún menor en los loci bialélicos y trialélicos ya que la frecuencia del alelo menor nunca baja del 25%.

3.2.4.2 Estima de la dosis alélica

Cada uno de los tres métodos anteriores ofreció una estima de la frecuencia a la que se encontraron los alelos detectados. En el caso de los alineamientos USEQs se utilizó el número de lecturas de las USEQs que se identificaron como alelos reales para calcular las frecuencias relativas de los distintos alelos. En el caso de AMPLIAN se utilizaron los valores de lecturas promedio que el programa asigna a cada haplotipo reconstruido. Por último, en el caso de AMPLISAS, se utilizó la profundidad de lecturas de cada alelo (suma de las lecturas idénticas al alelo real y de las que se correspondieron con otras secuencias asignadas al mismo grupo). Las frecuencias alélicas relativas se utilizaron para estimar la dosis alélica en los loci bialélicos (1:3, 2:2 o 3:1) y trialélicos (1:1:2, 1:2:1, 2:1:1) y para comparar con los valores esperados en los loci tetralélicos (1:1:1:1), ya que por la biología del parásito, no se esperan más de cuatro alelos por locus.

3.2.4.3 Análisis filogenético

El análisis filogenético se llevó a cabo utilizando MEGA versión 5 (Tamura *et al.*, 2011). Las distancias evolutivas entre alelos se calcularon siguiendo el modelo de Tamura-Nei (Tamura y Nei, 1993) y los árboles se construyeron utilizando el método Neighbor-Joining (Saitou y Nei, 1987). La fiabilidad de la topología resultante se evaluó mediante la técnica de múltiples réplicas o *bootstrap* ($N = 500$) (Felsenstein, 2009). Se utilizaron como grupos externos las secuencias del ensamblaje A (aislados WB y DH) disponibles en GiardiaDB: <http://giardiadb.org/giardiadb>.

También se incluyeron como referencia las secuencias de los genomas del ensamblaje B (aislados GS y BAH15), las secuencias de referencia de los subensamblajes BIII y BIV descritas en Wielinga y Thompson (2007) para *gdh* y *tpi*, las de los genotipos B1-B6 para *bg* descritas en Lalle *et al.* (2005), las del aislado BAH12c14 (Lasek-Nesselquist *et al.*, 2009) que se utilizó para definir el subensamblaje BIII en los ensayos de alozimas (Mayrhofer *et al.*, 1995) y las de los clones del experimento I (sin *singletons*).





4. Resultados



4 **Resultados**

4.1 **EXPERIMENTO I: ANÁLISIS DE LOS PATRONES DE DIVERSIDAD GENÉTICA DE *G. DUODENALIS* EN AISLADOS DE PACIENTES**

4.1.1 **Determinación de los ensamblajes y selección de las muestras**

Se seleccionó una muestra de 12 aislados representativos de los ensamblajes A y B de entre una colección recogida en el CHUS entre los años 2000 y 2010. Se determinó el ensamblaje de los aislados mediante amplificación específica del locus *tpi* hasta encontrar seis aislados de cada ensamblaje. Dos aislados que fueron positivos tanto para A como para B (122 y 321) se incluyeron en el análisis debido a la relevancia epidemiológica de las infecciones mixtas (suponen una oportunidad de recombinación entre ensamblajes, el efecto de las coinfecciones podría tener relevancia clínica,...). Debido a la menor variabilidad previamente reportada en el ensamblaje A (ver apartado 1.5.2) se incluyeron igualmente los seis aislados del ensamblaje A que fueron detectados (147, 152, 209, 251, 263, 839), y se completó la muestra con cuatro aislados del ensamblaje B, seleccionados entre aquellos que mejor amplificaron (407, 704, 1221, 1343). Las fechas de recogida de estos aislados se muestran en la Tabla 3.4.

4.1.2 **Amplificación, clonación y secuenciación de los loci *gdh*, *bg* y *calt***

Las amplificaciones de *gdh*, *bg* y *calt* con las parejas cebadores diseñados en las regiones conservadas de los ensamblajes A y B, permitieron la obtención de las secuencias de los ensamblajes esperados en la mayoría de los casos. Sin embargo, en algunas ocasiones no se produjo amplificación, se amplificó solo uno de los dos ensamblajes presentes en el aislado o se amplificaron regiones homólogas de otros organismos presentes en las heces. Esto obligó a utilizar cebadores más específicos y a confirmar mediante digestión con enzimas de restricción la identidad de los productos de amplificación. En la Tabla 4.1 se muestran las parejas de cebadores utilizadas para la amplificación de los tres loci en cada aislado.

Tabla 4.1: Diseño de las PCRs para la amplificación de los loci *gdh*, *bg* y *calt* en los aislados seleccionados.

PCR externa PCR interna Tamaño (pb)	<i>gdh</i>				<i>bg</i>				<i>calt</i>		
	F/R	F/R	F/R2	BF1/BR1	F/R	BF/BR	BF/R	F/R	F/R	F1/BR	F1/R1
	F2/R2	BF/BR	AF/AR	BF3/BR3	F1/R1	F/R	F/R1	BF1/R1	F1/R1	BF/BR	BF/BR
	494	405	434	398	415	421	419	323	364	315	315
Aislados	122	8-A			8-A				8-A	5-B	
	147	8-A	8-A		8-A				8-A		
	152	8*- <i>Bif</i>	6*	8-A	8-A				8-A		
	209	8-A	8-A		8-A				8-A		
	251	8-A			8-A				8-A		
	263			8-A	8-A				8-A		
	321	8-A	2-B		8-A			6-B/4-A	6-A/1-B		6-B
	407	8-B			8-B				8-B		
	704		8*- <i>Bif</i>	8-B			6-B				8-B
	839	8-A			8-A				8-A		
	1221	8-B			5-B	8-B			8-B		
	1342	8-B				8-B			8-B		

Nota: se indican las parejas de cebadores que fueron utilizadas en la amplificación de cada locus y aislado (con el nombre con el que se describen en la Tabla 3.1) y la longitud de los fragmentos obtenidos. Para cada amplificación se indican el número de secuencias clonales obtenidas y el ensamblaje al que fueron asignadas. Se señalan con un * los casos en que se obtuvieron secuencias pertenecientes a otros organismos. *Bif*: *Bifidobacterium*.

En dos aislados se obtuvieron secuencias homólogas del locus *gdh* de especies del género *Bifidobacterium*: en el aislado 152 se amplificó *Bifidobacterium longum* sp. *longum*, y en el 704 se amplificó *Bifidobacterium pseudocatenulatum*. El origen de estas secuencias fue determinado mediante un análisis de homología realizado con la herramienta BLAST (disponible en el portal del NCBI, *National Center for Biotechnology Information*; <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) en el que se obtuvieron valores de homología respecto a las secuencias de estos organismos del 100% y el 99%, respectivamente.

En las infecciones mixtas se utilizaron distintas combinaciones de cebadores específicos hasta encontrar las que amplificaron el ensamblaje B. Esto se debió a que los cebadores diseñados para amplificar específicamente este ensamblaje amplificaron en ocasiones el ensamblaje A. De hecho, en unas pruebas con cebadores diseñados para amplificar B, se obtuvieron productos de PCR en los aislados 147 y 209 (aislados A) que tras ser clonados y secuenciados confirmaron la amplificación inespecífica de A (Tabla 4.1). Finalmente, se obtuvieron secuencias B para los tres loci en el aislado 321, pero únicamente en *calt* en el aislado 122. El genotipado con *tpi* confirmó reiteradamente la presencia del ensamblaje B en este último aislado (comprobada mediante digestión con AluI, que corta el amplicón de *tpi* del ensamblaje B en dos fragmentos, pero no el del ensamblaje A).

En el aislado 321 se obtuvieron además dos secuencias de *calt* que fueron identificadas como recombinantes entre los ensamblajes A y B. Estas secuencias no se incluyeron en el análisis de diversidad ya que no pudieron ser asignadas a ningún ensamblaje ni se pudo descartar que se tratase de artefactos generados de manera espuria durante la PCR, como se discutirá más adelante.

4.1.3 Análisis de diversidad genética

4.1.3.1 Diversidad sinónima

La diversidad nucleotídica sinónima fue mayor en el ensamblaje B que en el A, tanto analizando el promedio de los tres loci ($4,14\% \pm 1,65\%$ vs. $0,77\% \pm 0,30\%$, respectivamente), como comparando cada uno de ellos por separado. *Gdh* presentó la mayor diferencia entre ensamblajes (π_s fue 29 veces mayor en B), siendo el más diverso de los tres en el ensamblaje B y el menos diverso en el ensamblaje A. *Calt* presentó también diferencias importantes entre ensamblajes, aunque mucho menores (π_s fue 4 veces mayor en B), y *bg*, por su parte, presentó los valores más similares entre ensamblajes (π_s apenas 1,5 veces mayor en B), siendo el menos diverso de los tres en B y sin embargo, el más diverso en A (Tabla 4.2).

Tabla 4.2: Diversidad genética en el conjunto de secuencias agrupadas por ensamblaje.

Ens	Locus	N	H	H_d	Posiciones sinónimas							Posiciones no sinónimas						
					L_S	S_S	π_s	$\pm EE$	θ_s	D_S	p	L_A	S_A	π_A	$\pm EE$	θ_A	D_A	p
A	<i>gdh</i>	80	12	0,50	81,8	2	0,25	0,22	0,49	-0,78		263,2	12	0,17	0,06	0,92	-2,25	**
	<i>bg</i>	67	21	0,69	67,3	7	1,29	0,75	2,18	-1,07		253,8	15	0,17	0,05	1,24	-2,51	***
	<i>calt</i>	62	6	0,48	63,5	3	0,77	0,68	1,01	-0,48		245,5	2	0,03	0,02	0,17	-1,44	
B	<i>gdh</i>	45	13	0,89	81,8	17	7,36	1,93	4,76	1,36		263,2	8	0,33	0,19	0,70	-1,42	
	<i>bg</i>	41	18	0,89	67,4	6	1,88	0,98	2,08	-0,30		253,6	12	0,23	0,06	1,11	-2,43	*
	<i>calt</i>	44	15	0,91	64,9	10	3,19	1,31	3,55	-0,33		250,2	5	0,07	0,04	0,46	-2,00	*

Nota: Ens: Ensamblaje; N: número de secuencias; H: número de haplotipos; H_d : diversidad haplotípica; L_S y L_A : número de posiciones sinónimas y no sinónimas; S_S y S_A : número de posiciones segregantes sinónimas y no sinónimas; π_s y π_A : diversidad nucleotídica sinónima y no sinónima expresada en porcentaje; θ_s y θ_A : diversidad nucleotídica basada en el número de posiciones segregantes sinónimas y no sinónimas expresada en porcentaje; D_S y D_A : D de Tajima en posiciones sinónimas y no sinónimas. EE: error estándar; p: significación estadística de la D de Tajima, *: $p < 0,05$, **: $p < 0,01$, ***: $p < 0,001$.

Tabla 4.3: Estimaciones de la diversidad genética por aislado en el ensamblaje A.

Locus	Aislado	N	H	H_d	Posiciones sinónimas							Posiciones no sinónimas						
					L_S	S_S	π_S	$\pm EE$	θ_S	D_S	p	L_A	S_A	π_A	$\pm EE$	θ_A	D_A	p
<i>gdh</i>	122	8	1	0,00	81,8	0	0,00	0,00	0,00			263,2	0	0,00	0,00	0,00		
	147	16	5	0,45	81,8	0	0,00	0,00	0,00			263,2	5	0,24	0,10	0,57	-1,93	*
	152	8	2	0,43	81,7	0	0,00	0,00	0,00			263,3	1	0,16	0,16	0,15	0,33	
	209	16	2	0,13	81,9	0	0,00	0,00	0,00			263,2	1	0,05	0,05	0,12	-1,16	
	251	8	2	0,25	81,8	0	0,00	0,00	0,00			263,2	1	0,10	0,10	0,15	-1,05	
	263	8	4	0,64	81,8	1	0,31	0,32	0,47	-1,05		263,2	4	0,38	0,19	0,59	-1,53	
	321	8	1	0,00	81,8	0	0,00	0,00	0,00			263,2	0	0,00	0,00	0,00		
	839	8	2	0,25	81,8	0	0,00	0,00	0,00			263,3	1	0,10	0,08	0,15	-1,05	
	Promedio		2,4	0,27		0,1	0,04		0,06				1,6	0,13		0,21		
	EE		0,5	0,08		0,1	0,04		0,06				0,7	0,05		0,08		
<i>bg</i>	122	8	2	0,25	67,4	0	0,00	0,00	0,00			253,6	1	0,10	0,10	0,15	-1,05	
	147	8	1	0,00	67,3	0	0,00	0,00	0,00			253,7	0	0,00	0,00	0,00		
	152	8	6	0,89	67,5	2	0,75	0,54	1,14	-1,31		253,5	3	0,28	0,21	0,46	-1,57	
	209	8	2	0,25	66,6	0	0,00	0,00	0,00			254,4	1	0,10	0,10	0,15	-1,05	
	251	8	4	0,75	67,1	2	1,64	1,18	1,15	1,45		253,9	2	0,10	0,10	0,30	-1,31	
	263	8	5	0,79	67,3	3	1,13	0,65	1,72	-1,45		253,8	2	0,20	0,14	0,30	-1,31	
	321	12	6	0,68	67,5	1	0,26	0,26	0,49	-1,14		253,5	5	0,33	0,17	0,65	-1,83	*
	839	7	4	0,81	67,3	3	1,89	1,12	1,82	0,05		253,7	1	0,11	0,11	0,16	-1,01	
	Promedio		3,8	0,55		1,4	0,71		0,79				1,9	0,15		0,27		
	EE		0,7	0,12		0,5	0,27		0,27				0,5	0,04		0,07		
<i>calt</i>	122	8	1	0,00	64,8	0	0,00	0,00	0,00			250,2	0	0,00	0,00	0,00		
	147	8	1	0,00	64,8	0	0,00	0,00	0,00			250,2	0	0,00	0,00	0,00		
	152	8	1	0,00	64,8	0	0,00	0,00	0,00			250,2	0	0,00	0,00	0,00		
	209	8	3	0,46	63,6	1	0,40	0,42	0,61	-1,05		245,4	1	0,10	0,98	0,16	-1,05	
	251	8	2	0,54	64,8	1	0,84	0,84	0,60	1,17		250,2	0	0,00	0,00	0,00		
	263	8	2	0,25	64,9	0	0,00	0,00	0,00			250,1	1	0,10	0,10	0,15	-1,05	
	321	6	2	0,33	64,8	1	0,52	0,54	0,68	-0,93		250,2	0	0,00	0,00	0,00		
	839	8	2	0,25	64,8	1	0,39	0,41	0,60	-1,05		250,2	0	0,00	0,00	0,00		
	Promedio		1,8	0,23		0,5	0,27		0,31				0,3	0,03		0,04		
	EE		0,3	0,08		0,2	0,11		0,12				0,2	0,02		0,03		

Nota: N : número de secuencias; H : número de haplotipos; H_d : diversidad haplotípica; L_S y L_A : número de posiciones sinónimas y no sinónimas; S_S y S_A : número de posiciones segregantes sinónimas y no sinónimas; π_S y π_A : diversidad nucleotídica sinónima y no sinónima expresada en porcentaje; θ_S y θ_A : diversidad nucleotídica basada en el número de posiciones segregantes sinónimas y no sinónimas expresada en porcentaje; D_S y D_A : D de Tajima en posiciones sinónimas y no sinónimas. EE: error estándar; p : significación estadística de la D de Tajima, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

Tabla 4.4: Estimaciones de la diversidad genética por aislado en el ensamblaje B.

Locus	Aislado	N	H	H _d	Posiciones sinónimas							Posiciones no sinónimas						
					L _S	S _S	π _S	±EE	θ _S	D _S	p	L _A	S _A	π _A	±EE	θ _A	D _A	p
<i>gdh</i>	122	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	321	13	6	0,64	81,5	1	0,19	0,18	0,40	-1,15		263,5	5	0,29	0,14	0,61	-1,58	
	407	8	3	0,68	81,8	6	3,97	1,64	2,83	1,60		263,2	1	0,22	0,21	0,15	1,44	
	704	8	2	0,25	81,8	0	0,00	0,00	0,00			263,2	1	0,10	0,10	0,15	-1,05	
	1221	8	1	0,00	81,8	0	0,00	0,00	0,00			263,2	0	0,00	0,00	0,00		
	1343	8	1	0,00	82,0	0	0,00	0,00	0,00			263,0	0	0,00	0,00	0,00		
	Promedio		2,6	0,31		1,4	0,83		0,6				1,4	0,12		0,18		
	EE		1,9	0,15		2,8	1,26		0,82				1,3	0,06		0,13		
<i>bg</i>	122	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	321	6	2	0,33	67,2	0	0,00	0,00	0,00			253,8	1	0,13	0,12	0,17	-0,93	
	407	8	4	0,75	67,4	2	1,29	0,98	1,14	0,41		253,6	1	0,10	0,10	0,15	-1,05	
	704	6	3	0,60	67,3	1	0,50	0,51	0,65	-0,93		253,7	1	0,13	0,13	0,17	-0,93	
	1221	13	7	0,83	67,5	2	1,39	0,97	0,96	1,21		253,6	4	0,24	0,12	0,51	-1,77	
	1343	8	6	0,89	67,4	1	0,37	0,37	0,57	-1,05		253,6	5	0,50	0,22	0,76	-1,60	
	Promedio		4,4	0,68		1,2	0,71		0,7				2,4	0,22		0,35		
	EE		1,9	0,10		0,6	0,22		0,21				1,4	0,05		0,13		
<i>calt</i>	122	5	2	0,40	64,9	0	0,00	0,00	0,00			250,1	1	0,16	0,16	0,19	-0,82	
	321	7	4	0,71	64,8	2	1,19	0,78	1,26	-0,27		250,2	2	0,23	0,16	0,33	-1,24	
	407	8	5	0,86	64,8	5	3,34	1,59	2,98	0,42		250,2	1	0,00	0,00	0,15	-1,05	
	704	8	1	0,00	64,8	0	0,00	0,00	0,00			250,2	0	0,00	0,00	0,00		
	1221	8	4	0,75	64,8	5	3,49	1,71	2,97	0,59		250,2	0	0,00	0,00	0,00		
	1343	8	2	0,25	64,9	0	0,00	0,00	0,00			250,1	1	0,10	0,11	0,15	-1,05	
	Promedio		3,0	0,50		2,0	1,34		1,2				0,8	0,08		0,14		
	EE		1,4	0,11		1,1	0,52		0,49				0,5	0,03		0,05		

Nota: Ver Tabla 4.3.

En todos los loci la diversidad sinónima del conjunto de secuencias agregadas de cada ensamblaje fue muy superior al promedio de las diversidades en cada uno de los aislados (Tablas 4.2, 4.3 y 4.4). El máximo exponente de esto fue el locus *gdh* en el ensamblaje B, en el que la diversidad agregada fue casi nueve veces superior a la del promedio de los aislados ($7,36\% \pm 1,93\%$ vs $0,83\% \pm 1,26\%$, Tablas 4.2 y 4.4 respectivamente). Esta observación puso de manifiesto que las diferencias genéticas entre aislados contribuyeron a incrementar de un modo importante la diversidad genética global (agregada) de ambos ensamblajes.

Los aislados mostraron niveles distintos de diversidad sinónima. En el ensamblaje B los aislados 1221 y 407 presentaron valores altos en dos y tres loci, respectivamente; sin embargo, 704 y 1343 resultaron prácticamente monomórficos (Tabla 4.4). En el ensamblaje A se reprodujo el mismo patrón, aunque a menor escala dada la baja variabilidad en este ensamblaje. Los aislados 251 y 839 mostraron los valores más elevados de diversidad y 122 y 147 no presentaron variantes sinónimas en ningún locus (Tabla 4.3).

El estadístico D de Tajima estimado para los lugares sinónimos (D_S) no reveló desviaciones significativas en el espectro de frecuencias de las variantes con respecto a lo esperado bajo la hipótesis de neutralidad (Tabla 4.2). Este patrón fue consistente tanto en los ensamblajes como en los aislados. En general, los valores de D_S de las secuencias agregadas fueron negativos en A y en B, lo que sugirió un ligero exceso de variantes a baja frecuencia. Esto se reflejó en la presencia de 20 mutaciones que aparecieron una única vez en la muestra (*singletons*). El locus *gdh*, sin embargo, mostró un valor positivo en el ensamblaje B, que además fue el que más se desvió de la neutralidad en términos absolutos, aunque no llegó a mostrar niveles estadísticamente significativos.

4.1.3.2 Diversidad no sinónima

La diversidad no sinónima fue mucho menor que la sinónima en los tres loci tanto en A como en B (π_A agregada promedio: 0,12% \pm 0,05% y 0,21% \pm 0,07%, respectivamente; Tabla 4.2). En este caso, el promedio de la diversidad dentro de los aislados (Tablas 4.3 y 4.4) fue similar al obtenido en el conjunto de cada ensamblaje (secuencias agrupadas, Tabla 4.2). Las diferencias entre aislados también fueron menores, sin que ninguno de ellos presentase un valor de π_A que destacase sobre los demás (Tablas 4.2, 4.3 y 4.4).

Los valores de θ_A fueron superiores a los de π_A , lo que implica un exceso de variantes a baja frecuencia en las posiciones no sinónimas respecto a lo esperado bajo el equilibrio mutación-deriva. Como cabía esperar, esto se reflejó en los valores de D_A , que fueron negativos tanto en A como en B en los tres loci, alcanzando desviaciones estadísticamente significativas en *gdh* y *bg* en A, y en *bg* y *calt* en B (Tablas 4.3 y 4.4). La D_A promedio de los tres loci también mostró valores negativos y estadísticamente significativos para ambos ensamblajes ($D_A = -3,60$; $p < 0,001$ y $D_A = -3,42$; $p < 0,001$ para A y B, respectivamente).

Esto es indicativo de que las posiciones no sinónimas se encuentran bajo el efecto de fuerzas evolutivas distintas a las que actúan sobre las posiciones sinónimas.

4.1.3.3 Análisis de la varianza molecular

Para estudiar en mayor profundidad la estructura de la diversidad genética en la muestra, se llevó a cabo un análisis de los componentes de la covarianza. Se investigó la contribución de las diferencias entre aislados y dentro de aislados a la varianza total y se observó que tanto en el ensamblaje A como en el B, ambas resultaron significativas. En el ensamblaje A la varianza dentro de aislados fue ligeramente superior (promediando los tres loci un $53,33\% \pm 8,34\%$, Tabla 4.5), mientras que en el ensamblaje B fueron las diferencias entre aislados las que contribuyeron en mayor medida a la varianza total ($59,47\% \pm 15,13\%$; Tabla 4.5). En este sentido destacó especialmente *gdh*-B, donde las diferencias entre aislados representaron el 86,58% de la varianza total. En cambio, en *bg*-B tan solo representaron un 34,27%.

Tabla 4.5: Análisis de varianza molecular.

Ensamblaje	Locus	Fuente de varianza	g.l.	Suma de cuadrados	Varianza	% de varianza	<i>p</i>
A	<i>gdh</i>	Entre aislados	7	12,90	0,17	45,82	< 0,0001
		Dentro de aislados	72	14,25	0,20	54,18	< 0,0001
		Total	79	27,14	0,37		
	<i>bg</i>	Entre aislados	7	16,95	0,23	32,66	< 0,0001
		Dentro de aislados	59	28,32	0,48	67,34	< 0,0001
		Total	66	45,27	0,71		
	<i>calt</i>	Entre aislados	7	18,19	0,31	61,51	< 0,0001
		Dentro de aislados	54	10,49	0,19	38,49	< 0,0001
		Total	61	28,69	0,55		
B	<i>gdh</i>	Entre aislados	4	142,67	3,94	86,58	< 0,0001
		Dentro de aislados	40	24,45	0,61	13,42	< 0,0001
		Total	44	167,12	4,55		
	<i>bg</i>	Entre aislados	4	17,35	0,44	34,27	< 0,0001
		Dentro de aislados	36	30,20	0,84	65,73	< 0,0001
		Total	40	47,55	1,28		
	<i>calt</i>	Entre aislados	5	31,51	0,78	57,55	< 0,0001
		Dentro de aislados	38	21,98	0,58	42,45	< 0,0001
		Total	43	53,48	1,36		

Nota: g.l., grados de libertad; *p* nivel de significación (probabilidad de obtener el resultado observado en una muestra en la que el porcentaje de ese componente de la varianza es igual a cero).

4.1.3.4 Diversidad haplotípica

Ambos ensamblajes mostraron valores elevados de diversidad haplotípica, y al igual que ocurrió con la diversidad nucleotídica, H_d fue mayor en el ensamblaje B (promedio de los tres loci: $0,89 \pm 0,01$ en B vs. $0,56 \pm 0,07$ en A, Tabla 4.2). Este parámetro depende tanto del número de haplotipos como de sus frecuencias, pero no tiene en cuenta el número de posiciones variantes entre haplotipos, de modo que solo está parcialmente relacionado con las estimas de diversidad nucleotídica; por ejemplo, en el ensamblaje B los loci *bg* y *gdh* presentaron la misma diversidad haplotípica (Tabla 4.2), mientras que la diversidad nucleotídica total fue tres veces mayor en *gdh* (1,90% vs. 0,57%).

Los haplotipos observados en cada locus, así como el número de veces que fueron detectados en cada muestra aparecen representados en las Tablas 8.1-8.6 (Anexos). Para la descripción de los haplotipos se utilizó el alineamiento de las secuencias completas, incluyendo las regiones no solapantes que habían sido excluidas del análisis de diversidad. Por una parte esto permitió describir las variantes que se encontraban en los extremos de las secuencias más largas, pero también generó un número reducido de nuevos haplotipos que solo se diferenciaron en la longitud de la secuencia a pesar de ser idénticos en la región solapante (señalados con asteriscos en las Tablas 8.1-8.6). En total se describieron 119 haplotipos distintos, de los cuales tan solo 34 se secuenciaron en más de un clon.

El patrón de distribución de los haplotipos fue radicalmente distinto en los ensamblajes A y B. En A existe un haplotipo mayoritario en cada uno de los loci, que está presente en prácticamente todas las muestras: hA2 en *gdh* (hA3 y hA4 son probablemente versiones más cortas de este mismo haplotipo), hA1 en *bg* y hA1 en *calt* (Tablas 8.1, 8.3 y 8.5) seguidos por uno o dos haplotipos que aparecieron a alta frecuencia en un aislado (hA1 en *gdh*) o un número reducido de aislados (hA10 en *gdh*, hA11 en *bg* y hA2 en *calt*). El resto de haplotipos se correspondieron con *singletons* (secuencias que aparecen una única vez en la muestra).

El ensamblaje B presentó un patrón distinto, en el que el mismo haplotipo rara vez apareció en más de un aislado. En sentido estricto esto solo ocurrió con hB4 de *calt* (aislados 321, 497 y 1221), aunque es razonable pensar que podrían existir otros cuatro casos en los que haplotipos de distintos aislados fueron iguales, ya que se diferenciaron únicamente en la longitud de las secuencias, uno en *calt*: hB1 (122) y hB11 (321); y tres en *bg*: hB5 (407) y

hB19 (1221), hB1 (321) y hB3 o hB5 (407), y hB1 (321) y hB19 (1221) (Tablas 8.4 y 8.6). En todo caso se trató de eventos aislados en el contexto del ensamblaje B, donde la mayor parte de los aislados presentaron un haplotipo o un conjunto de haplotipos característicos y distintos a los observados en los demás aislados. Además, en concordancia con la mayor diversidad genética en este ensamblaje, los haplotipos presentaron más cambios entre sí que los del ensamblaje A. Estas diferencias entre haplotipos fueron especialmente llamativas en el locus *gdh*, en el que cada uno de los haplotipos principales (excluyendo los *singletons*) presentó al menos tres sustituciones nucleotídicas respecto a cualquier otro (Tablas 8.1-8.6).

4.1.3.5 Análisis filogenético

En los tres loci las secuencias se dividieron en dos ramas principales, formando sendos grupos alrededor de las secuencias de referencia de A y B, respectivamente. Estos grupos presentaron un soporte estadístico superior al 99% y la distancia genética entre ellos fue muy superior a las observadas entre las secuencias dentro de cada grupo (Figuras 4.1, 4.2 y 4.3). Estos datos indicaron que los dos ensamblajes representan entidades genéticas diferenciadas.

Dentro de los ensamblajes se observó una estructura mucho menos definida y los distintos subgrupos presentaron valores de *bootstrap* generalmente bajos. Las secuencias de *gdh* y *calt* del ensamblaje A se agruparon junto a la secuencia de referencia del subensamblaje AII y separadas de la del subensamblaje AI. La escasa variación en la región analizada de *bg* impidió separar las secuencias de referencia de los subensamblajes AI y AII. Sin embargo, este locus fue el único en que un subgrupo de A presentó valores de *bootstrap* superiores al 75%. Fue el grupo formado por hA11 con el genotipo A3 (Lalle et al. 2005), que se diferenciaron en dos SNVs de la secuencia de referencia.

En el ensamblaje B se distinguieron dos grupos principales en *gdh*: uno incluyó las secuencias de referencia del subensamblaje BIV y similares y el otro las del subensamblaje BIII (*bootstrap* 76%, Figura 4.1). El haplotipo hB8, mayoritario en el aislado 407, se situó en una posición divergente anterior a la separación de BIII y BIV. Esta división no se observó en los demás loci. En *gdh*, los haplotipos descritos en los aislados 321, 1221 y 1343 formaron grupos propios con valores de *bootstrap* > 80%, mostrando una vez más que la diversidad detectada en este locus se encuentra estructurada en aislados (Figura 4.1).

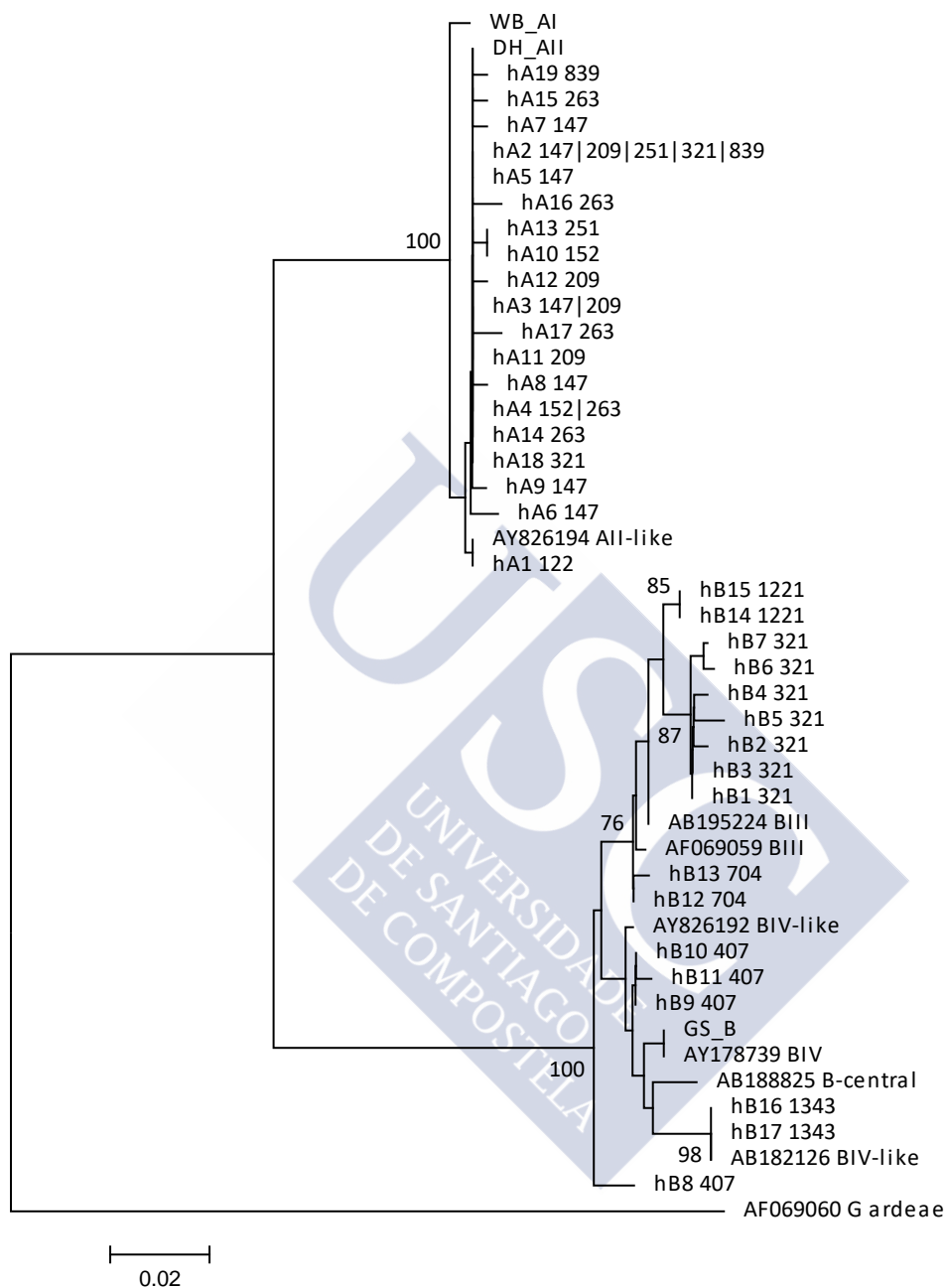
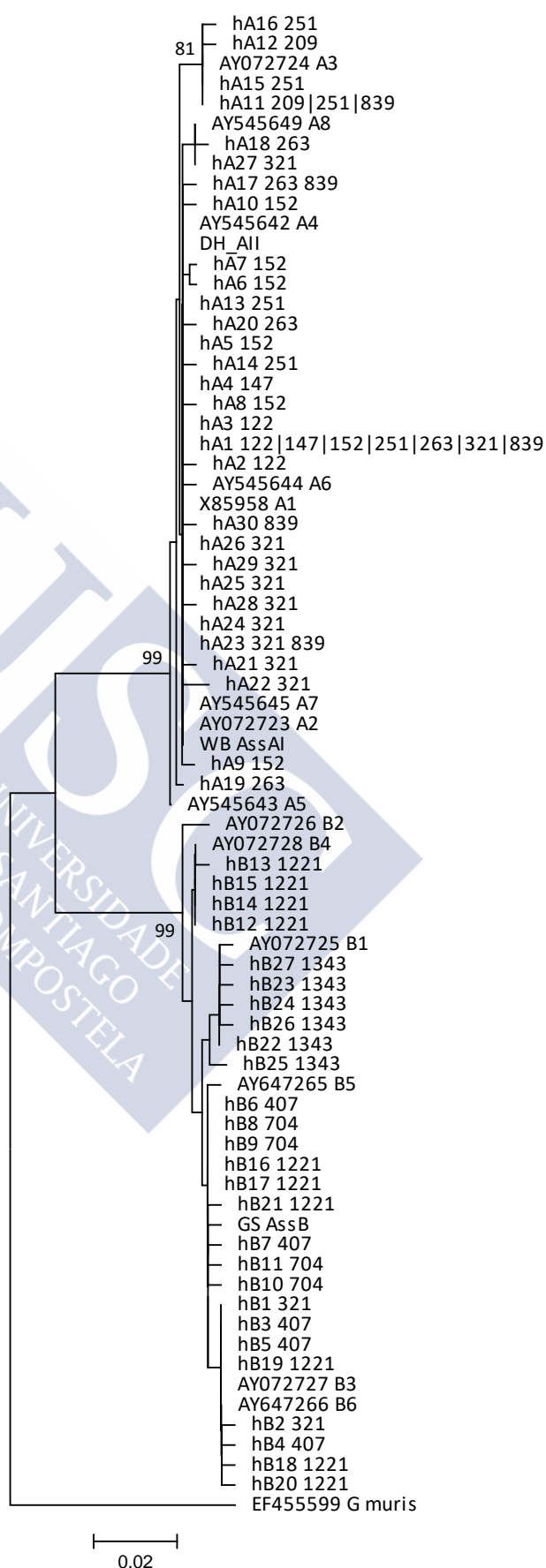


Figura 4.1: Árbol filogenético de los haplotipos del locus *gdh*. El nombre de cada taxón indica el haplotipo (como en las Tablas 8.1 y 8.4) y los aislados en los que aparece separados por barras verticales. En el caso de las secuencias de referencia se indica el número de acceso de GenBank y el subensamblaje/subtipo según Wieling y Thompson (2007). También se incluyen las secuencias de los genomas de referencia de los aislados GS (B), WB (AI) y DH (AII). Los valores de bootstrap superiores a 75 se muestran al lado de los nodos.

Figura 4.2: Árbol filogenético de los haplotipos del locus *bg*. El nombre de cada taxón indica los haplotipos iguales en la región solapante (como en las Tablas 8.2 y 8.5) separados por guiones y los aislados en los que aparecen separados por barras verticales. En el caso de las secuencias de referencia se indica el número de acceso de GenBank y el subtipo según Lalle *et al.* (2005). También se incluyen las secuencias de los genomas de referencia de los aislados GS (B), WB (AI) y DH (AII). Los valores de bootstrap superiores a 75 se muestran al lado de los nodos.



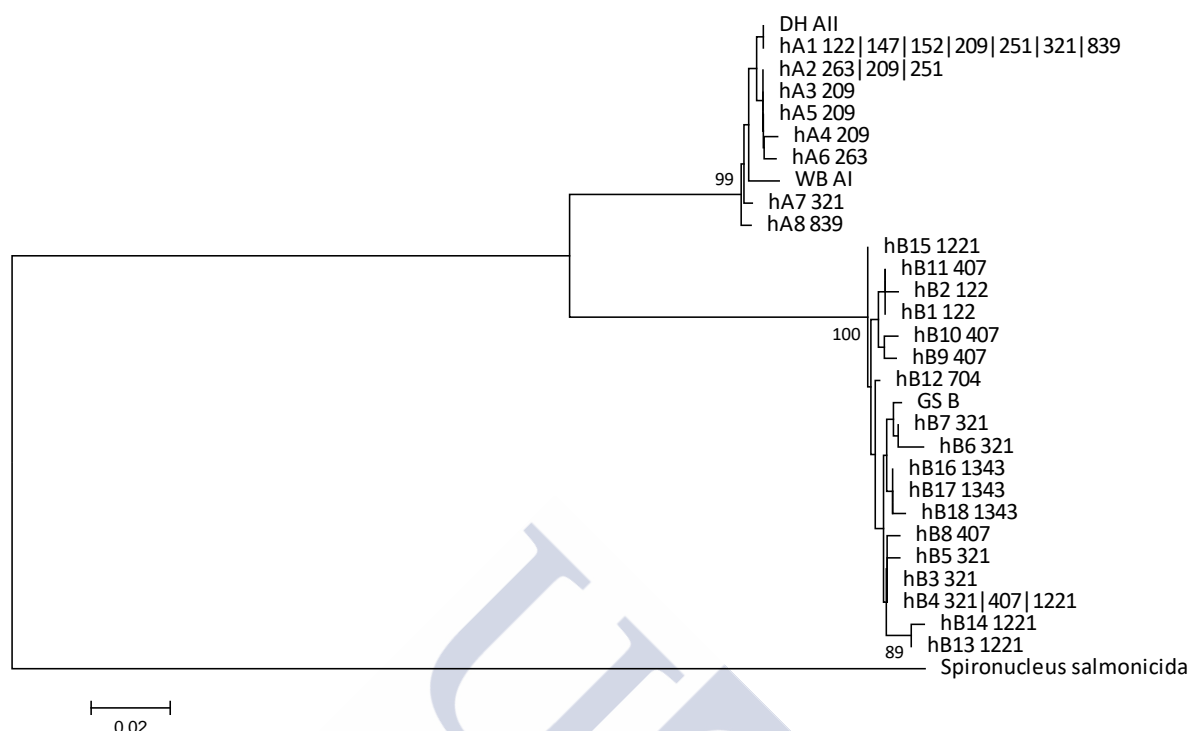


Figura 4.3: Árbol filogenético de los haplotipos del locus *calt*. El nombre de cada taxón indica el haplotipo (como en las Tablas 8.3 y 8.6 y los aislados en los que aparece separadas por barras verticales. También se incluyen las secuencias de los genomas de referencia de los aislados GS (B), WB (AI) y DH (AII). Los valores de bootstrap superiores a 75 se muestran al lado de los nodos.

Las diferencias en los patrones de diversidad nucleotídica y haplotípica de los ensamblajes A y B también se vieron reflejadas en la estructura de los árboles. La baja diversidad observada en el ensamblaje A dio como resultado una estructura filogenética simple en la que la distancia entre haplotipos fue pequeña y en la que apenas se formaron subgrupos. En concreto, se distinguieron el subgrupo de *bg* mencionado anteriormente, un subgrupo en *calt* formado por secuencias de los aislados 152, 209 y 263 (que se distinguieron de las demás por una única mutación), y dos subgrupos en *gdh*: el que formaron los haplotipos hA1 característico del aislado 122 junto a la secuencia de referencia AII-like (AY826194), y el que formaron los haplotipos hA10 y hA13, que eran idénticos en la región solapante. En ambos casos, estos grupos se diferenciaron del genoma de referencia (WB) en una única mutación (Figuras 4.1, 4.2 y 4.3).

En el caso del ensamblaje B, la elevada diversidad provocó que los haplotipos formasen más subgrupos y que a su vez existiesen mayores distancias genéticas entre ellos. Aun así,

muy pocos grupos obtuvieron un soporte de *bootstrap* por encima del 75%. Cabe destacar que en la filogenia de *gdh* cada aislado formó un subgrupo específico, con la excepción de los haplotipos del aislado 407 que aparecieron divididos en dos clados situados en posiciones alejadas del árbol. Sin embargo, en las filogenias de *bg* y *calt* la separación de los aislados en subgrupos no fue una constante. Las secuencias del aislado 1343 formaron subgrupos independientes en ambos casos, mientras que las de 407 y 1221 aparecieron en distintas posiciones de los árboles, intercaladas entre ellas o con otras secuencias de los aislados 321 y 704 (Figuras 4.1, 4.2, 4.3).

4.1.3.6 Redes de haplotipos

Para comprobar si la población de *G. duodenalis* de Santiago de Compostela (Galicia) se diferencia genéticamente respecto a las poblaciones de otras áreas geográficas, se realizó un análisis de redes de haplotipos incluyendo las secuencias de *gdh* y *bg* y otras disponibles en GenBank de origen humano y procedencia geográfica conocida (secuencias solapantes, sin bases degeneradas). La mayor parte de los haplotipos de los ensamblajes A y B observados en los aislados del área sanitaria de Santiago de Compostela resultaron idénticos o muy similares a otros haplotipos descritos previamente en localizaciones tan distantes como Asia, África o América.

Las redes del ensamblaje A mostraron una estructura sencilla (Figura 4.4). En *gdh* se observó un haplotipo principal correspondiente con la secuencia del genoma de referencia AII, rodeado de haplotipos que se diferenciaron en una única mutación, y un segundo haplotipo menos frecuente correspondiente con la secuencia del genoma de referencia AI. En *bg*, se observó un patrón similar pero con dos haplotipos comunes presentes en un número similar de aislados, rodeados de haplotipos que se diferenciaron de éstos en una o dos posiciones. Las redes de haplotipos del ensamblaje B presentaron una estructura mucho más compleja y reticulada, sobre todo en el caso de *gdh* (Figura 4.5). Tanto en *gdh* como en *bg* se observaron un mayor número de haplotipos comunes –presentes en aislados de distintos continentes, incluyendo los de este trabajo– relacionados entre sí por intrincadas redes filogenéticas. Las redes reticuladas cerradas, sugieren que algunos haplotipos podrían haberse generado por recombinación y no por mutación, ya que implicarían que una misma posición ha mutado en dos veces.

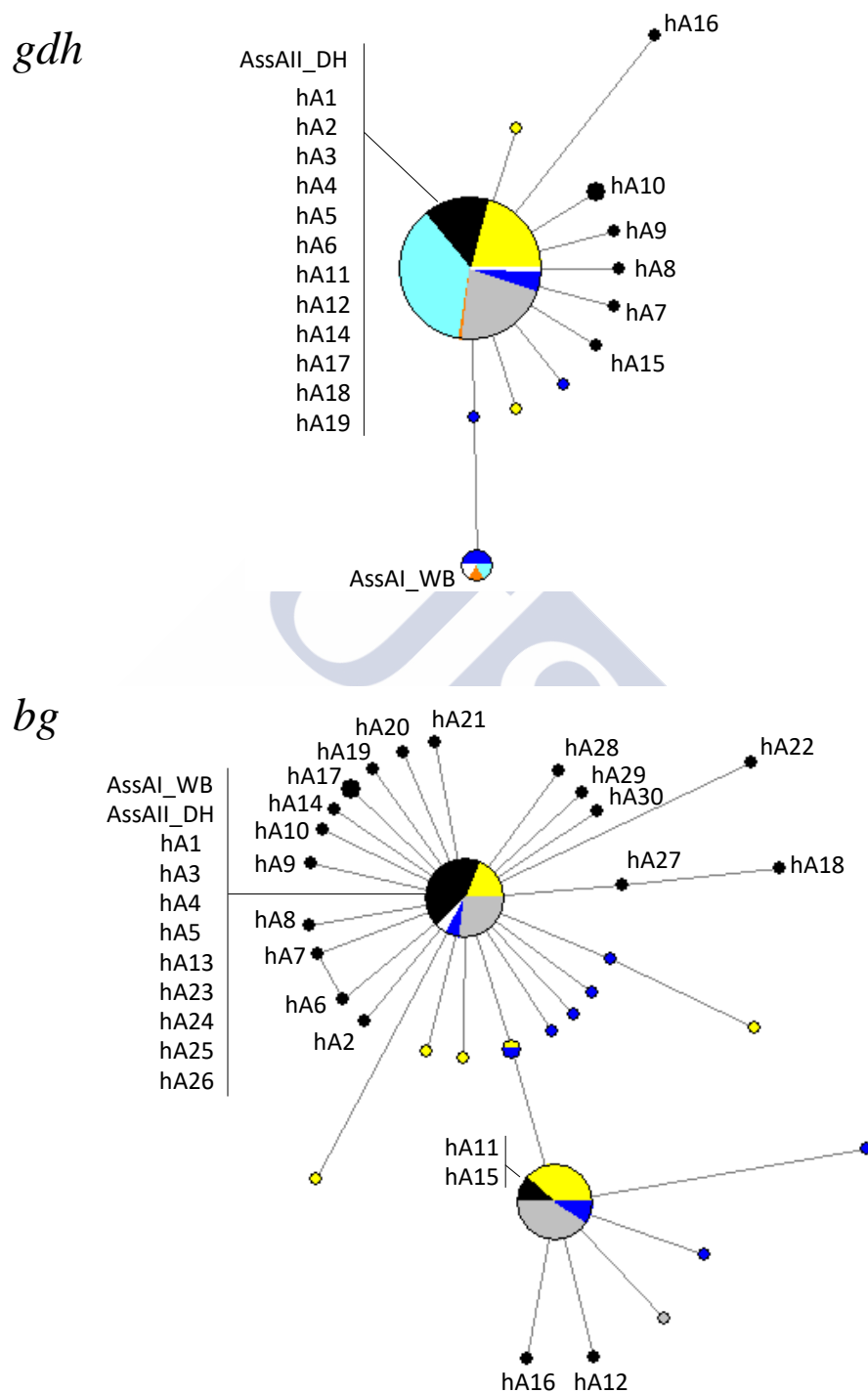


Figura 4.4: Redes de haplotipos del ensamblaje B de los loci *gdh* y *bg*. Los haplotipos están representados por círculos cuyo tamaño es proporcional al número de aislados en que se detectaron. El código de colores identifica el origen geográfico de las muestras. Negro: este estudio; blanco: genomas de referencia; azul claro: América; azul oscuro: Europa; gris: África; amarillo: Asia.

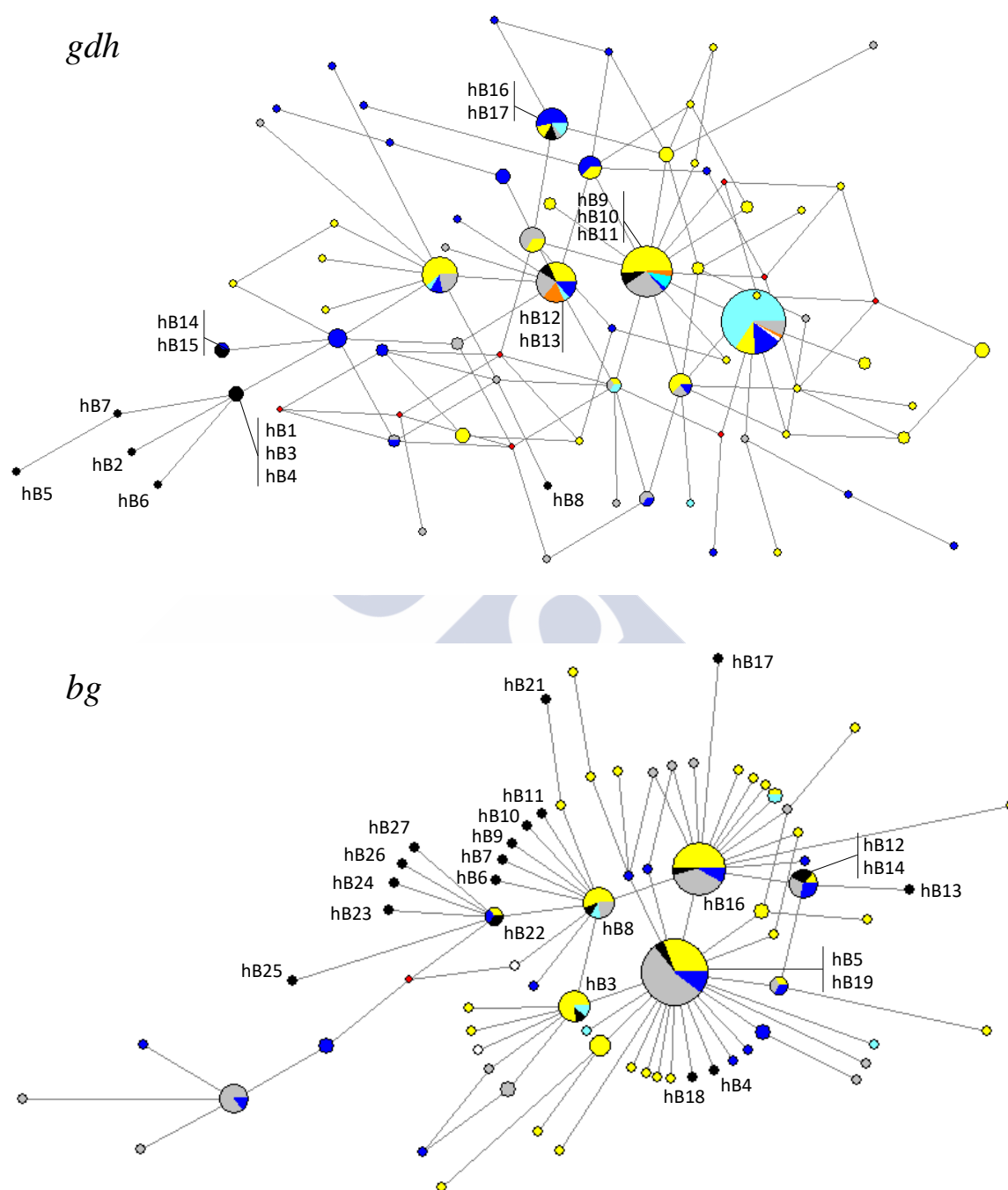


Figura 4.5: Redes de haplotipos del ensamblaje A de los loci *gdh* y *bg*. Los haplotipos están representados por círculos cuyo tamaño es proporcional al número de muestras en que se detectaron. El código de colores identifica el origen geográfico de las muestras. Negro: este estudio; blanco: genomas de referencia; azul claro: América; azul oscuro: Europa; gris: África; amarillo: Asia; naranja: Australia. Los nodos rojos representan secuencias intermedias hipotéticas que no se encontraron en los alineamientos.

4.1.3.7 Recombinación

Dados los patrones de evolución reticulada observados en las redes de haplotipos en el ensamblaje B, se estudió la presencia de recombinación en los tres loci dentro de cada ensamblaje y en el conjunto de la muestra. Ninguna de las pruebas utilizadas detectó evidencias significativas de recombinación dentro o entre ensamblajes. Las estimas de la tasa de recombinación a escala poblacional fueron muy bajas o iguales a cero (Tabla 4.6). El valor más elevado se detectó en *calt* dentro del ensamblaje A, pero los análisis de permutación demostraron que ninguno de los valores era significativamente distinto de cero. El estadístico R_h , que estima el límite inferior del número de eventos de recombinación necesarios para explicar los haplotipos observados en la muestra bajo un modelo de posiciones infinitas (ninguna posición muta más de una vez), fue igual a cero en los tres loci para el ensamblaje A y mostró valores bajos tanto en el ensamblaje B como en el conjunto de la muestra.

Tabla 4.6: Análisis de recombinación.

Locus	<i>N</i>	<i>L</i>	R_h	ρ	<i>Lkmax</i>	<i>p</i>	<i>G4</i>	<i>p</i>	$\text{corr}(r^2, d)$	<i>p</i>	$\text{corr}(D' , d)$	<i>p</i>
<i>gdh</i>	125	346	6	0	-201 702,9	0,28	5543	0,36	-0,06	0,11	-0,01	0,33
<i>gdh B</i>	45	346	2	0	-12 472,0	0,46	255	0,69	0,03	0,62	0,04	0,69
<i>gdh A</i>	80	346	0	1	-2329,7	0,70	0	1,00	-0,02	0,41	-	1,00
<i>bg</i>	108	323	5	0	-90 250,0	0,39	2422	0,50	0,00	0,50	0,00	0,51
<i>bg B</i>	41	323	0	0	-3645,0	0,75	0	1,00	0,03	0,65	-	1,00
<i>bg A</i>	67	323	0	6	-4738,0	0,09	0	1,00	-0,01	0,46	-	1,00
<i>calt</i>	106	315	4	0	-107 922,0	0,83	4083	0,61	0,05	0,82	0,04	0,64
<i>calt B</i>	44	315	2	7	-3291,0	0,31	150	0,68	-0,20	0,01	0,69	0,73
<i>calt A</i>	62	315	0	12	-344,0	0,27	0	1,00	-0,16	0,20	-	1,00

Nota: *N*: número de secuencias; *L*: longitud del alineamiento; R_h : límite inferior del mínimo número de eventos de recombinación; ρ : estima de la tasa de recombinación a escala poblacional $4N_e r$; *Lkmax*: máxima verosimilitud; *G4*: suma de la distancia entre todos los pares de posiciones segregantes para los que se detectan las cuatro combinaciones haploides posibles; $\text{corr}(r^2, d)$: correlación entre la estima de desequilibrio de ligamiento (DL) r^2 y la distancia física; $\text{corr}(|D'|, d)$: correlación entre la estima de DL $|D'|$ y la distancia física; *p*: significación estadística en los test de permutación (1000 permutaciones).

4.1.3.8 Divergencia

La divergencia promedio entre los ensamblajes A y B en las posiciones sinónimas fue del $71,88\% \pm 0,22\%$. *Calt* presentó la mayor diferenciación entre ensamblajes en las posiciones sinónimas, con valores de K_S superiores a la unidad aplicando la corrección de Jukes Cantor (Tabla 4.7). Esto implica que algunas de las posiciones sinónimas en *calt* mutaron más de una vez desde que los dos ensamblajes comenzaron a diferenciarse. La divergencia en posiciones no sinónimas fue más de un orden de magnitud inferior en todos los casos, determinando una proporción K_A/K_S próxima a cero en todos los loci.

Tabla 4.7: Divergencia entre ensamblajes.

Locus	$K_S(JC)$	$K_A(JC)$	K_A/K_S
<i>gdh</i>	58,63%	2,22%	0,0380
<i>bg</i>	41,57%	0,20%	0,0049
<i>calt</i>	115,43%	0,06%	0,0005

Nota: JC: corrección de Jukes-Cantor

4.2 EXPERIMENTO II: GENOTIPADO MULTILOCUS DE QUISTES INDIVIDUALES DE *G. DUODENALIS* MEDIANTE SECUENCIACIÓN MASIVA EN PARALELO

4.2.1 Purificación de quistes

Para aislar los quistes mediante micromanipulación fue necesario obtener una suspensión de quistes libre de material fecal. Para ello se utilizó un gradiente de sacarosa. La menor densidad de los quistes respecto a la disolución de sacarosa permitió recuperarlos de la interfase sin restos de residuos, que precipitaron al fondo (Figura 4.6).

Se seleccionaron aislados frescos que presentaron una elevada concentración de quistes y una buena apariencia externa, medida en términos de refringencia y silueta. En los aislados que presentaban quistes rotos o poco refringentes se observó un elevado número de quistes en el fondo del tubo en lugar de en la interfase, probablemente debido a la pérdida de sus características densiométricas al romperse.

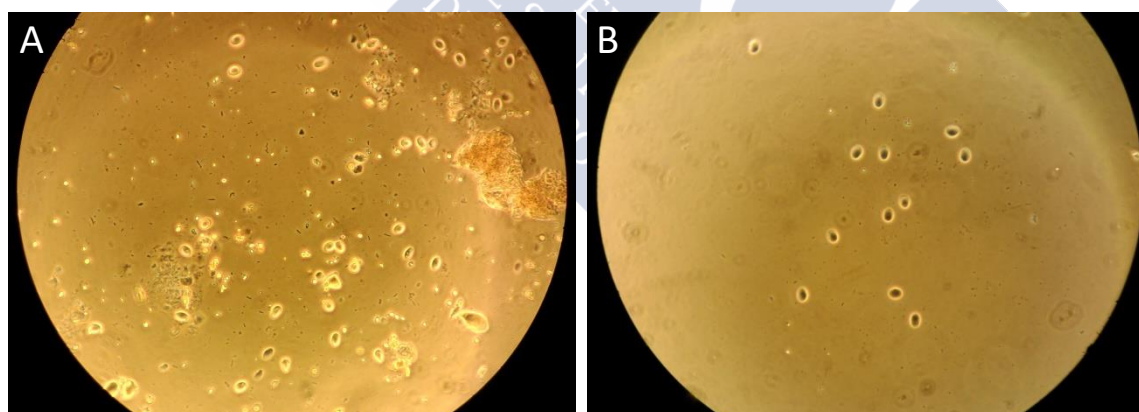


Figura 4.6: Imágenes de microscopia de contraste de fase de la suspensión de quistes. A) Después de los lavados iniciales con solución salina. B) Después del gradiente de sacarosa. Los quistes se observaron a 400 aumentos.

Los aislados utilizados en la puesta a punto de los protocolos y los que fueron incluidos en el experimento de secuenciación se nombraron por orden alfabético, según llegaron al laboratorio. Finalmente, se secuenciaron los quistes de los aislados D, E, H, K y M. Los

quistes individuales recibieron el nombre del aislado seguido por un número arábigo, siguiendo el orden por el que fueron separados.

4.2.2 Amplificación multiplex de quistes aislados

4.2.2.1 Optimización de la PCR multiplex

En primer lugar se amplificó cada locus por separado utilizando 1 µL de la suspensión de quistes purificados como molde (>100 quistes) y se realizaron pruebas a diferentes temperaturas de anillamiento. A 63°C, se obtuvieron bandas únicas en cada uno de los loci a intensidades similares.

En segundo lugar, se realizaron las pruebas de amplificación de los siete loci (*bg*, *tpi*, *rpL7a*, *mlh1*, *calt*, *fd* y *gdh*) en multiplex, ajustando la concentración de cada pareja de cebadores hasta obtener una intensidad similar en todas las bandas. La amplificación multiplex produjo un patrón de cinco bandas distinguibles en los geles de agarosa (2,5% p/v), ya que los amplicones de *mlh1* y *rpL7a*, así como los de *fd* y *calt*, eran de tamaños muy similares y no llegaron a separarse.

Finalmente, se realizaron las amplificaciones de quistes individuales. En la mayor parte de los quistes, no se produjo amplificación del locus *tpi*, por lo que fue reamplificado y añadido posteriormente a la multiplex para obtener una representación similar a la del resto de loci (tal y como se indica en Material y Métodos).

4.2.2.2 Control de la amplificación de ADN libre

La extracción de conclusiones correctas a partir de los resultados obtenidos con este diseño experimental exige que las secuencias provengan de un único quiste. Para ello, fue necesario asegurar la presencia de un único quiste en cada PCR y la ausencia de ADN libre en suspensión –resultado de la liberación del contenido citoplasmático de otros quistes–.

Con el propósito de eliminar el ADN en suspensión se realizaron distintas pruebas. La primera consistió en la realización de hasta cinco lavados por centrifugación tras el gradiente de sacarosa. Después de cada uno de los lavados se reservó una alícuota de sobrenadante con

el objetivo de comprobar si existía una disminución en la amplificación a medida que se aumentaba el número de lavados. En lugar de lavar los quistes, las centrifugaciones promovieron la amplificación de ADN libre (Figura 4.7), lo que invalidó esta aproximación.

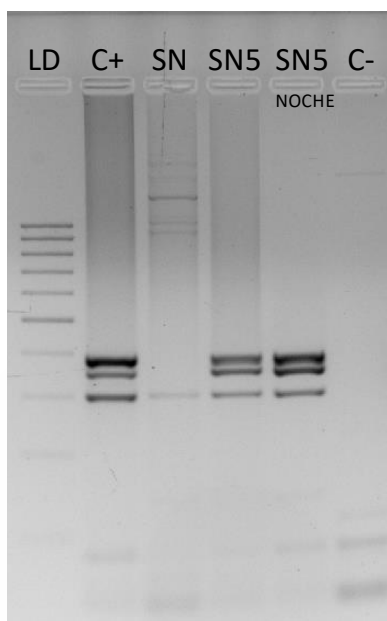


Figura 4.7: Electroforesis de la amplificación del sobrenadante (SN) de la suspensión de quistes. Como molde para la amplificación de ADN se utilizó: C+: 1 μ L de la suspensión de quistes; SN: 1 μ L de SN recogido tras el último lavado de la purificación de sacarosa; SN5: 1 μ L de SN recogido tras cinco lavados por centrifugación; SN5 noche: 1 μ L de SN recogido tras cinco lavados por centrifugación después de guardar los quistes toda la noche a 4°C; C-: 1 μ L agua.

A tenor de estos resultados las centrifugaciones fueron reducidas al mínimo necesario para lavar el exceso de sacarosa que se arrastra cuando los quistes se recuperan de la interfase (una o, como máximo, dos). Además, los quistes se purificaron justo antes de la micromanipulación, es decir, no se conservaron purificaciones de quistes de un día para otro.

En segundo lugar, se realizaron pruebas de digestión del ADN libre con DNaseI para tratar de eliminar cualquier resto de ADN que no estuviese protegido dentro de los quistes. Sin embargo, la acción de esta enzima no solo eliminó el ADN libre, sino que inhibió la amplificación, incluso cuando se utilizó como molde 1 μ L de la suspensión de quistes (Figura 4.8). Por esta razón este método también fue descartado para el tratamiento de los quistes individuales.

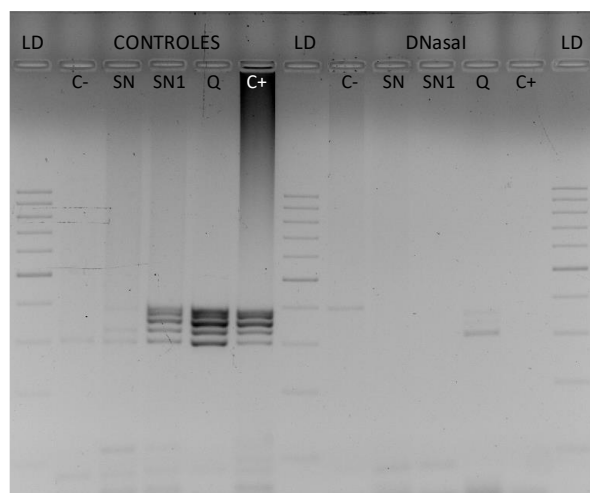


Figura 4.8: Electroforesis de la amplificación de muestras tratadas y sin tratar con DNasaI. Los controles recibieron el mismo tratamiento que las muestras añadiendo agua en lugar de la enzima. C-: control negativo, 1 μ L de agua; SN: 1 μ L de sobrenadante recogido tras el último lavado de la purificación con sacarosa; SN1: 1 μ L de sobrenadante recogido tras guardar los quistes toda la noche a 4°C; Q: 1 μ L de suspensión de quistes; C+: control positivo, 1 μ L producto de PCR diluido 1:1000.

En tercer lugar, se decidió incluir una serie de lavados y diluciones durante la micromanipulación. Antes de iniciar la micromanipulación se dejó sedimentar los quistes en el tubo al que habían sido transferidos tras el gradiente de sacarosa, se retiró el sobrenadante y se sustituyó por 10 mL de PBS limpio. De este modo los posibles restos de ADN libre se diluyeron antes de la micromanipulación. En la preparación de la muestra para la micromanipulación se aplicó una nueva dilución: se depositaron 400 μ L de la suspensión de quistes en una placa de Petri que se cubrió con PBS, lo que diluyó el sobrenadante sin afectar a la concentración de quistes por unidad de superficie, ya que éstos precipitan rápidamente.

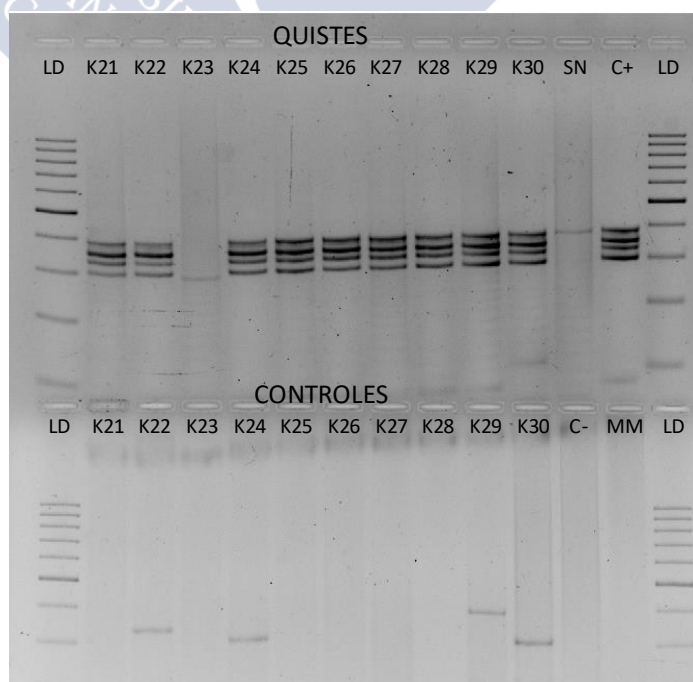
También se añadieron dos lavados tras el aislamiento de los quistes: cada quiste se depositó en sendas gotas de 400 μ L de PBS, se lavó por repipeteo y se recogió de nuevo. Esto supuso que el aislamiento de los quistes se comprobó por triplicado, lo que descarta la posibilidad de que durante la manipulación se arrastrase más de un quiste. Por último se establecieron controles negativos para cada uno de los quistes, y en los casos en que apareció una banda a la altura esperada de un amplicón, el quiste correspondiente fue descartado.

4.2.2.3 PCR directa multiplex de quistes individuales

Las reacciones de amplificación se realizaron a continuación de la micromanipulación sin aplicar procesos de ruptura de la pared, ni extracción de ADN. La tasa de amplificación de los quistes individuales utilizando esta metodología fue de 60/75 (80%; aunque la amplificación del locus *tpi* fue menos productiva y tuvo que ser reamplificado). Ensayos previos demostraron que el uso de quistes conservados a -20°C , o sometidos a un proceso de desnaturalización (5 min a 95°C seguidos de 15 min a -80°C) resultaba en tasas de amplificación mucho más bajas.

A pesar de que en los geles de agarosa se cargó el volumen completo de las PCRs de los controles (en contraste con las reacciones a partir de quistes de las que *solo* se cargaron 2 μL), en ningún caso un control negativo presentó un patrón de bandas comparable con el de los quistes. Aparecieron en cambio bandas aisladas más tenues, similares a las obtenidas en los quistes que fallaron (*e.g.*, quiste K23, Figura 4.9). Los controles negativos de 12 de los 60 quistes en los que se obtuvieron productos de amplificación, mostraron alguna banda entre 300 y 400 pb, por lo que los quistes correspondientes fueron directamente descartados (*e.g.*, quistes K22, K24, K29, K30, Figura 4.9).

Figura 4.9: Electroforesis de la PCR multiplex de quistes individuales de *G. duodenalis* y sus controles de ADN libre. En la fila superior se cargaron 2 μL de los productos de amplificación de: 10 quistes aislados (K21-K30), un control negativo (SN) y un control de la muestra con >100 quistes (C+). En la fila inferior se cargaron los 10 μL de los controles del sobrenadante de cada quiste y otros dos controles negativos (C-: 1 μL de agua sobre un fragmento de vidrio como los utilizados para depositar los quistes; y MM: 1 μL de agua añadido a la mezcla de PCR). Los quistes K22, K24, K29 y K30 que presentaron alguna banda en los controles negativos fueron descartados.



4.2.3 Resultados de la secuenciación masiva en paralelo

4.2.3.1 Rendimiento de la reacción de secuenciación

En la secuenciación masiva en paralelo se incluyeron ocho quistes y un control positivo (>100 quistes) de cinco aislados (D, E, H, K y M). Una vez filtradas las lecturas de baja calidad, las ISP policlonales y los fragmentos control, se obtuvieron 2 001 552 lecturas independientes utilizables con una longitud promedio de 211 pb, lo que supone más de 400 Mb secuenciadas (Figura 4.10). El rendimiento de la reacción de secuenciación estuvo por debajo de lo esperado debido a dos razones principales: i) la eficiencia de carga del chip fue muy baja respecto a lo habitual en este tipo de experimentos, lo que provocó que la mitad de los micropocillos disponibles en el chip de secuenciación se encontrasen vacíos; ii) entre los pocillos que contenían ISP, un 50% presentaron ISP policlonales, un valor que se encuentra por encima de lo esperado.

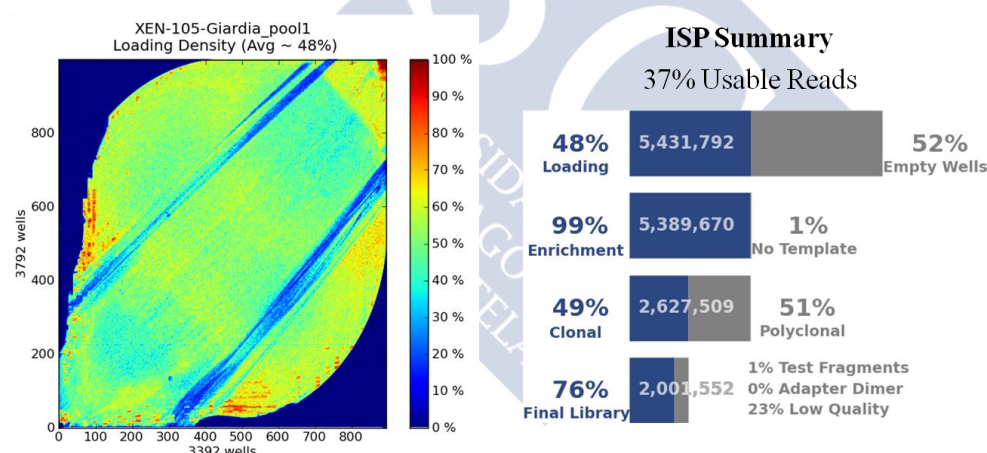


Figura 4.10. Rendimiento de la carrera de secuenciación. A) Gráfico de densidad de carga de ISP en el chip de secuenciación 318v2. B) Datos del informe de secuenciación: % de pocillos con ISP, % de ISP con ADN molde, % de ISP clonales. % de lecturas finales.

Del total de lecturas utilizables, 341 147 no fueron asignadas a ningún código de barras y 78 006 pertenecieron a tres muestras de otro experimento que fueron secuenciadas en la misma carrera de secuenciación. Finalmente un total de 1 582 374 lecturas fueron alineadas contra el genoma de *G. duodenalis*.

El promedio de lecturas totales por quiste/muestra fue de $29\,229,96 \pm 1998,45$ de las cuales un 57% fueron lecturas completas (longitud superior al 95% del amplicón). En el cálculo de esta media no se tuvieron en cuenta las lecturas que se correspondieron únicamente con la secuencia de los cebadores y que no llegaron a alinear con al menos una posición interna del amplicón.

El promedio de lecturas por quiste varió ampliamente entre loci, desde algo menos de 300 en *calt* hasta más de 10 000 en *bg*. Estos valores reflejaron grandes diferencias entre loci en la tasa de amplificación durante la PCR multiplex que no fueron detectados previamente a la secuenciación. Sin embargo, la reamplificación del locus *tpi* dio buenos resultados, siendo el segundo locus más representado con un promedio de más de 8000 lecturas por quiste. Por otra parte, cabe destacar que la proporción de lecturas que cubrieron la totalidad del amplicón varió en función del locus, alcanzando un máximo del 70,5% en *rpl7a*, y un mínimo de apenas el 27,8% en *fd*. Curiosamente, estos dos valores coincidieron con la mayor y la menor tasa de *indels*, respectivamente (Tabla 4.8).

Tabla 4.8: Valores promedio de cobertura, número de lecturas y tasa de *indels* por quiste.

Locus	<i>bg</i>	<i>tpi</i>	<i>RPL7</i>	<i>mlh1</i>	<i>calt</i>	<i>fd</i>	<i>gdh</i>
Cobertura promedio	7249,1 <i>566,1</i>	7103,6 <i>710,1</i>	1920,2 <i>154,2</i>	768,0 <i>94,6</i>	218,2 <i>32,7</i>	282,5 <i>33,0</i>	4225,3 <i>377,6</i>
Q30	4130,0 <i>327,0</i>	4007,2 <i>400,1</i>	1159,6 <i>94,2</i>	423,0 <i>52,4</i>	117,4 <i>17,7</i>	131,5 <i>15,4</i>	2216,7 <i>194,6</i>
Lecturas	10656,3 <i>750,8</i>	8265,2 <i>822,8</i>	2242,9 <i>176,5</i>	1118,7 <i>127,7</i>	295,3 <i>45,8</i>	484,3 <i>54,3</i>	6167,2 <i>529,9</i>
MAPQ > 60	8874,1 <i>689,9</i>	7815,1 <i>779,6</i>	2167,5 <i>170,9</i>	833,6 <i>98,9</i>	256,8 <i>39,1</i>	365,4 <i>42,0</i>	5256,6 <i>435,6</i>
Longitud > 95%	6112,4 <i>494,0</i>	5824,1 <i>592,8</i>	1704,6 <i>139,4</i>	484,2 <i>61,6</i>	145,5 <i>23,2</i>	134,6 <i>16,2</i>	2168,9 <i>210,1</i>
USEQs	1629,1 <i>111,7</i>	2222,5 <i>211,2</i>	623,9 <i>43,6</i>	254,0 <i>27,4</i>	87,8 <i>12,3</i>	120,9 <i>13,7</i>	1397,3 <i>116,6</i>
>1 lectura	240,5 <i>17,0</i>	259,4 <i>25,0</i>	82,6 <i>6,6</i>	28,7 <i>3,4</i>	7,4 <i>1,2</i>	4,7 <i>0,7</i>	107,4 <i>10,2</i>
Tasa de <i>indels</i> (%)	0,46% <i>0,01%</i>	0,50% <i>0,02%</i>	0,38% <i>0,01%</i>	0,46% <i>0,01%</i>	0,50% <i>0,02%</i>	0,89% <i>0,02%</i>	0,67% <i>0,02%</i>

Nota: en cursiva se indica el error estándar de cada promedio. Q30: cobertura con *Phred-like quality score* > 30. MAPQ: *mapping quality*, identidad > 60% respecto al genoma de referencia; Longitud > 95%: lecturas que cubren > 95% del amplicón. USEQs: secuencias únicas; > 1: secuencias únicas con más de una lectura. Tasa de *indels* (%): promedio de *indels* por cada 100 pb. Los valores de cobertura no incluyen las posiciones de los cebadores.

El número de lecturas también varió ampliamente entre quistes. Algunos llegaron a presentar hasta siete veces más lecturas que otros (E13 vs. M22, Figura 4.11), a pesar de que las librerías fueron cuantificadas y mezcladas a una concentración equimolar.

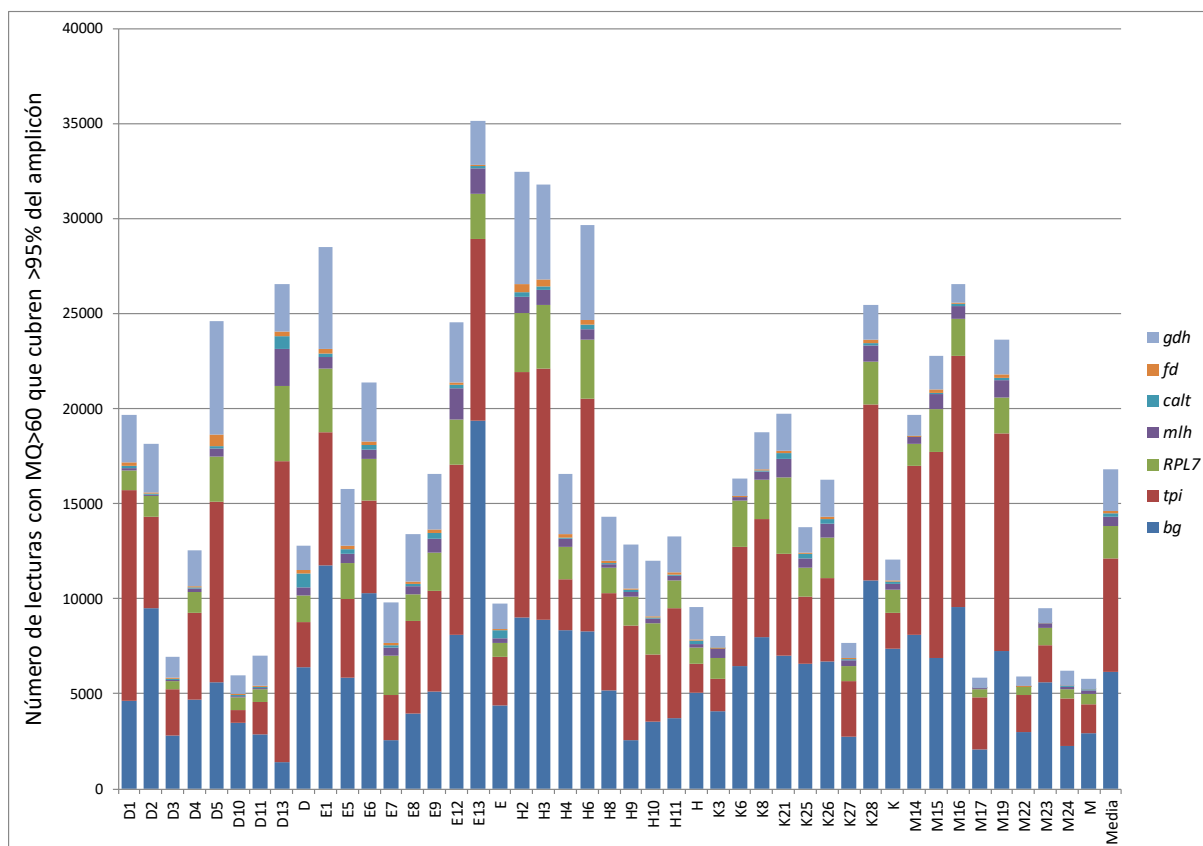


Figura 4.11: Número de lecturas por amplificación de cada quiste y control de cada aislado (negrita). El total de lecturas se encuentra dividido por locus según el código de colores especificado en la leyenda.

4.2.4 Genotipado multilocus

La variación genética en un individuo de *Giardia* se puede estructurar en varios niveles: i) la heterozigosidad alélica, ii) la dosis alélica, es decir, el número de copias de cada alelo, iii) la distribución de los alelos en los núcleos y iv) la asociación de los alelos en las fases haploides. Sin embargo, el diseño del experimento no permitió inferir cómo se distribuyeron los alelos en los cuatro núcleos de cada quiste. Por tanto, los genotipos multilocus obtenidos

en este trabajo, se limitan a una descripción de los alelos de los distintos loci presentes en cada individuo, sin que se conozca cómo se asocian dentro de cada uno de los núcleos.

4.2.4.1 Identificación de alelos

La identificación de los alelos reales presentes en cada quiste se llevó a cabo utilizando tres métodos: i) USEQ, un método de filtrado de secuencias únicas que se obtuvieron al colapsar las lecturas idénticas, ii) AMPLIAN, un método probabilístico de corrección de errores e inferencia de haplotipos y iii) AMPLISAS, una herramienta *online* de análisis de genotipado mediante agrupamiento de secuencias por homología (ver Material y Métodos).

En cada quiste se identificaron entre uno y cuatro alelos por amplicón, en concordancia con lo esperado dada la ploidía y el ciclo vital del organismo (máximo cuatro alelos distintos, ver Ciclo Vital). En total, se identificaron siete alelos de *mlh1*, seis de *bg*, *rpL7a*, *calt* y *fd* y cinco de *tpi* y *gdh*. Los SNVs que definen cada uno de los alelos, así como el número de quistes en los que fueron detectados se muestran en las Tablas 8.7-8.13 (Anexos).

AMPLIAN fue el método que permitió identificar alelos en un mayor número de quistes: 311 de los 315 amplicones analizados (7 loci x 45 quistes/aislados). Los alineamientos de secuencias únicas USEQ presentaron problemas en los loci con menor cobertura, ya que en 23 amplicones no se obtuvo ninguna USEQ que cumpliera los criterios para poder ser considerada un alelo real (estar representada por más de una lectura, cubrir el amplicón por completo y no presentar *indels* que alterasen la pauta de lectura). Por último, AMPLISAS establece una profundidad de lectura mínima por amplicón, es decir, un valor mínimo de lecturas por debajo de las cuales se considera que no hay suficientes lecturas para producir un resultado fiable por este método. En base a las observaciones realizadas en los alineamientos de USEQs, este parámetro se estableció en 50 lecturas, lo que conllevó la pérdida de resultados en *mlh1*, *calt* y *fd* en 5, 24 y 32 quistes, respectivamente (Tablas 8.14-8.18, Anexos).

En resumen, en 252 de 315 amplicones se obtuvieron resultados por los tres métodos, en tres de ellos ninguno de los métodos propuso alelos reales (*calt* en los quistes M22 y M24 y *fd* en el quiste D3), en otros 20 fallaron USEQ y AMPLISAS y en 40 falló alguno de los tres métodos: 1 USEQ, 1 AMPLIAN, 38 AMPLISAS (Tablas 8.14-8.18, Anexos).

En los 252 casos en los que los tres métodos identificaron alelos, los resultados coincidieron en 230, lo que supuso una concordancia entre métodos del 91% y puso de manifiesto la robustez del protocolo establecido para el genotipado. En los casos en que se obtuvieron resultados de USEQ y AMPLIAN, pero no en AMPLISAS por falta de cobertura (profundidad de lectura < 50), la concordancia fue del 84% (32/38).

La causa más común de discrepancias entre métodos fue la pérdida de alelos en USEQ. Este método es el más sencillo de los tres, y por tanto el que más sufre el efecto de los artefactos de secuenciación. Por ejemplo, este método fue más proclive a descartar alelos potencialmente reales en aquellos casos en los que la presencia de *indels* redujo el número de secuencias utilizables. Este problema puede afectar de forma diferencial a distintos alelos ya que los *indels* son dependientes de secuencia (ver más abajo). Esto no sucedió en la misma medida con AMPLIAN y AMPLISAS porque utilizan algoritmos más sofisticados que corrigen estos errores de secuenciación lo que les permite identificar estos alelos sin problemas.

En los amplicones donde los tres métodos identificaron alelos, el promedio de lecturas asignadas a cada alelo por USEQ, AMPLIAN y AMPLISAS fue de $475,6 \pm 55,2$; $848,0 \pm 88,6$ y $1266,3 \pm 118,5$; respectivamente. Los alelos detectados para cada amplicón, así como el número de lecturas asignadas por cada método, se muestran en las Tablas 8.14-8.18 (Anexos)

4.2.4.1.1 Identificación de artefactos

El método USEQ permitió monitorizar el efecto de los filtros sobre el conjunto real de secuencias únicas (USEQs), identificando cuántas se descartaron en cada paso y a cuántas lecturas correspondían (Figura 4.12).

Esto convirtió a USEQ en el método más adecuado de los tres para identificar los errores introducidos durante la secuenciación o en la PCR. La inspección manual de los alineamientos USEQ después de cada paso también permitió ajustar los parámetros utilizados en el filtrado. Así mismo, ayudó en la interpretación de los resultados de AMPLIAN y AMPLISAS, ya que supuso un punto de comparación entre las lecturas manipuladas por los algoritmos de estos métodos (corregidas o agrupadas) y las lecturas reales que generó el secuenciador.

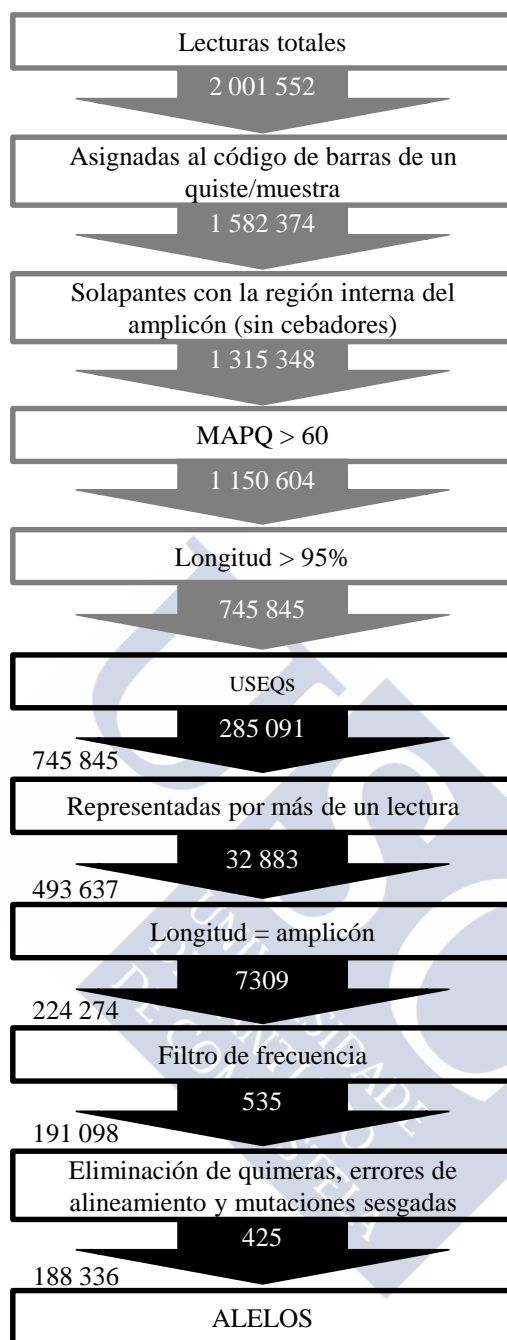


Figura 4.12: Esquema del filtrado de lecturas. En las flechas grises se indican el número de lecturas totales que pasaron cada filtro hasta ser agrupadas en USEQs. En las flechas negras se indica el número de USEQs que pasaron cada filtro hasta que se identificaron los alelos reales. Al lado de estas flechas se indica el número de lecturas representadas por las USEQs que pasaron cada filtro.

En las Tablas 4.9, 4.10 y 4.11 se muestran las secuencias detectadas en el locus *mlh1* en el quiste K21 por los métodos USEQ, AMPLIAN y AMPLISAS, respectivamente, antes de eliminar

los *indels*, las quimeras, los errores de alineamiento y las variantes sesgadas en el sentido de las lecturas. En estos alineamientos se pueden observar los artefactos más comunes y cómo se distribuyen alrededor de dos haplotipos más frecuentes (*mlh1_3* y *mlh1_4*) que se identificaron como los alelos reales. Los 2 alelos fueron las secuencias con mayor número de lecturas (244 lecturas, un 48,1% del total), seguidas por las 39 secuencias que presentaron *indels* (230 lecturas, 45,4%) y las 9 secuencias recombinantes (33 lecturas, 6,5%) (Tabla 4.9).

Este patrón se repitió de modo general en todos los alineamientos. Los *indels* fueron el error más común, tanto en términos de número de errores (número de posiciones distintas con *indels*) como en términos de frecuencia (número de lecturas con un *indel* en una posición concreta). Ocasionalmente algunas secuencias con *indels* alcanzaron frecuencias similares e incluso superiores a las de los correspondientes alelos en pauta de lectura con la secuencia de referencia. Este tipo de artefactos, conocidos como *highly frequent indels* (HFI), son característicos de esta tecnología de secuenciación y se producen habitualmente durante la secuenciación de homopolímeros de tres o más nucleótidos (Bragg *et al.*, 2013). Se trata de errores sistemáticos que se producen al secuenciar una cadena de nucleótidos concreta en un orden concreto, lo que facilitó su identificación ya que aparecieron en un único sentido de las lecturas. Por estas razones y dado que no se espera que existan mutaciones que alteren la pauta de lectura en los genes analizados, se descartaron las secuencias de longitud distinta a la esperada, lo que simplificó notablemente los alineamientos y facilitó la identificación de los alelos reales. En el caso de la región intrónica de *rpL7a*, cada *indel* fue comprobado individualmente y descartado en base al sesgo en el sentido de las lecturas.

A diferencia de los *indels*, las mutaciones puntuales, las quimeras y las combinaciones de *indels* que se cancelan (mismo número de inserciones que delecciones) no alteran la longitud de las secuencias y por ello tuvieron que ser identificados y descartados manualmente en los casos en que superaron el filtro de frecuencia: 7,5% de las lecturas del mismo tamaño que el amplicón. Esto ocurrió tanto en USEQ como en AMPLIAN y AMPLISAS, donde este tipo de errores también se detectaron habitualmente (Tablas 4.9, 4.10 y 4.11).

Tabla 4.9: Alineamiento USEQ del locus *mlh1* del quiste K21

USEQs	Lecturas	Posiciones variantes																													
		217	222	226	226	247	250	254	274	291	292	299	299	303	306	337	340	359	363	371	379	404	420	426	430	440	446	449	460	537	554
Ens B_aislado GS		T	C	-	T	A	T	A	T	C	-	-	T	A	T	T	A	C	-	T	A	-	A	G	-	A	-	T	T	-	-
mlh1_3	81
mlh1_3-222del	44	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
mlh1_3-222del-340del	36	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
mlh1_3-340del	16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
mlh1_3-337del	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
mlh1_3-299del	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
mlh1_3-225_226insT	4	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
mlh1_3-222del-337del-340del	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
mlh1_3-298_299insT	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
mlh1_3-553_554insT	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
mlh1_3-222del-340del-403_404insA	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
mlh1_3-222del-226del-340del	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
mlh1_3-222del-226del	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
mlh1_3-225_226insT-340del	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
mlh1_3-429_430insT	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
mlh1_3-445_446insT-553_554insT	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
mlh1_3-445_446insT	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
mlh1_3-460del	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
mlh1_3-303del-536_537insT	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
mlh1_3-274del	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
mlh1_4	163	.	.	.	G	.	G	.	T	-	-	-	G	C	.	G	-	-	-	G	C	-	G	-	C	-	-	-	-	-	-
mlh1_4-222del	20	-	-	-	G	.	G	.	T	-	-	-	G	C	.	G	-	-	-	G	C	-	G	-	C	-	-	-	-	-	-
mlh1_4-340del	13	-	-	-	G	.	G	.	T	-	-	-	G	C	.	G	-	-	-	G	C	-	G	-	C	-	-	-	-	-	-
mlh1_4-553_554insT	8	-	-	-	G	.	G	.	T	-	-	-	G	C	.	G	-	-	-	G	C	-	G	-	C	-	-	-	-	-	-
mlh1_4-337del	8	-	-	-	G	.	G	.	T	-	-	-	G	C	.	G	-	-	-	G	C	-	G	-	C	-	-	-	-	-	-
mlh1_4-254del	7	-	-	-	G	.	-	-	T	-	-	-	G	C	.	G	-	-	-	G	C	-	G	-	C	-	-	-	-	-	-
mlh1_4-222del-340del	3	-	-	-	G	.	G	.	T	-	-	-	G	C	.	G	-	-	-	G	C	-	G	-	C	-	-	-	-	-	-
mlh1_4-217del	6	-	-	-	G	.	G	.	T	-	-	-	G	C	.	G	-	-	-	G	C	-	G	-	C	-	-	-	-	-	-
mlh1_4-337del-340del	3	-	-	-	G	.	G	.	T	-	-	-	G	C	.	G	-	-	-	G	C	-	G	-	C	-	-	-	-	-	-
mlh1_4-250del	3	-	-	-	G	.	-	-	T	-	-	-	G	C	.	G	-	-	-	G	C	-	G	-	C	-	-	-	-	-	-
mlh1_4-254del-553_554insT	2	-	-	-	G	.	-	-	T	-	-	-	G	C	.	G	-	-	-	G	C	-	G	-	C	-	-	-	-	-	-
mlh1_4-362_363insT-553_554insT	2	-	-	-	G	.	G	.	T	-	-	-	G	C	.	G	-	-	-	G	C	-	G	-	C	-	-	-	-	-	-
mlh1_4-536_537insT	2	-	-	-	G	.	G	.	T	-	-	-	G	C	.	G	-	-	-	G	C	-	G	-	C	-	-	-	-	-	-
mlh1_4-254del-291_292insC	2	-	-	-	G	.	-	-	T	-	-	-	G	C	.	G	-	-	-	G	C	-	G	-	C	-	-	-	-	-	-
mlh1_4-379del	3	-	-	-	G	.	G	.	T	-	-	-	G	C	.	G	-	-	-	G	C	-	G	-	C	-	-	-	-	-	-
mlh1_4-371del	2	-	-	-	G	.	G	.	T	-	-	-	G	C	.	G	-	-	-	G	C	-	G	-	C	-	-	-	-	-	-
mlh1_4-217del-379del	2	-	-	-	G	.	G	.	T	-	-	-	G	C	.	G	-	-	-	G	C	-	G	-	C	-	-	-	-	-	-
mlh1_4-217del-222del	2	-	-	-	G	.	G	.	T	-	-	-	G	C	.	G	-	-	-	G	C	-	G	-	C	-	-	-	-	-	-
mlh1_4-247del-254del	2	-	-	-	-	-	-	-	T	-	-	-	G	C	.	G	-	-	-	G	C	-	G	-	C	-	-	-	-	-	-
quimera	7	-	-	-	G	.	G	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
quimera	6	-	-	-	G	.	G	.	T	-	-	-	G	C	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
quimera	4	-	-	-	G	.	G	.	T	-	-	-	G	C	.	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
quimera	3	-	-	-	-	-	-	-	T	-	-	-	G	C	.	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
quimera	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
quimera	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
quimera	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
quimera	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
quimera-217del	2	-	-	-	G	.	G	.	T	-	-	-	G	C	.	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
quimera-254del-553_554insT	2	-	-	-	G	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Nota: Se usó la secuencia del aislado GS como referencia. Los puntos indican posiciones conservadas. Los guiones indican huecos en el alineamiento. Amarillo: *indels*. Gris: región procedente del alelo *mlh1_3* (límites arbitrarios establecidos entre posiciones informativas).

Tabla 4.10: Alineamiento de las secuencias inferidas por amplian en *mlh1* en el quiste K21 ($p>0.95$).

Haplotipos	Prob.	Lecturas	Posiciones variantes																											
			236	237	247	254	291	299	303	304	305	306	337	339	340	353	355	358	359	378	379	420	426	427	428	434	438	439	440	449
Ens B_aislado GS			G	C	A	A	C	T	A	T	C	T	T	C	A	A	C	G	C	G	A	A	G	G	A	T	C	A	A	T
mlh1_3	1.00	268.5
mlh1_3–340del	1.00	103.7
mlh1_3–299del	1.00	27.7
mlh1_3–337del–340del	1.00	15.0
mlh1_4	1.00	381.6	.	.	G	G	T	.	G	.	.	C	G	.	.	G	C	G	C
mlh1_4–378G>A–379A>G	1.00	10.2	.	.	G	G	T	.	G	.	.	C	G	A	G	G	C	G	C	.	.
mlh1_4–254del	1.00	15.0	.	.	G	–	T	.	G	.	.	C	G	.	.	G	C	G	C	.	.
quimera	1.00	23.6	.	.	G	G
quimera	1.00	32.6	.	.	G	G	T	.	G	.	.	C
quimera	1.00	23.5	T	.	G	.	.	C	G	.	.	G	C	G	C	.	.
quimera	1.00	19.2	G	.	.	G	C	G	C	.	.
quimera	1.00	12.1	G	C	G	C	.	.
quimera–254del	1.00	8.1	.	.	G	–
quimera –254del	1.00	4.7	.	.	G	–	T	.	G	.	.	C
quimera–múltiples errores	1.00	7.4	.	.	–	–	.	.	T	C	G	.	.	G	C	G	C	.	.
Múltiples errores	1.00	7.3	C	T	C
Múltiples errores	1.00	7.0	A	T	C	–	–	C
Múltiples errores	0.99	4.9	A	T	C	–	G	C
Múltiples errores	0.99	4.0	A	C	C	G	C	G	G	.	.	.
Múltiples errores	0.97	2.2	A	G	.	G	T	.	G	.	.	–

Nota: Ver Tabla 4.9.

Tabla 4.11: Alineamiento de las secuencias identificadas por amplias en *mlh1* en el quistes K21.

		Posiciones variantes									
Alelos	Lecturas	247	254	291	303	306	359	420	426	440	449
Ens B_aislado GS		A	A	C	A	T	C	A	G	A	T
<i>mlh1_3</i>	187
<i>mlh1_4</i>	269	G	G	T	G	C	G	G	C	G	C
<i>quimera</i>	18	G	G
<i>quimera</i>	15	G	G	T	G	C

Nota: Ver Tabla 4.9.

Las secuencias con mutaciones puntuales fueron mucho menos frecuentes que las secuencias con *indels* y no alcanzaron la frecuencia suficiente para ser consideradas variantes reales, por lo que fueron descartadas. Aquellas que superaron el umbral de frecuencia fueron comprobadas en el IGV y descartadas si presentaban sesgo en el sentido de las lecturas. Tan solo dos variantes de este tipo fueron consideradas reales: las que definen los alelos *tpi_5* y *gdh_5* (Tablas 8.8 y 8.13, respectivamente, Anexos).

Por su parte, las quimeras representaron el tipo de secuencia más común después de los alelos reales y los HFI. Las secuencias identificadas como recombinantes potenciales de dos alelos más comunes en el mismo amplicón se consideraron artefactos y se descartaron sobre la base de las siguientes evidencias:

- i) Estas secuencias potencialmente recombinantes se detectaron en todos los alineamientos en los que se identificó más de un alelo real, y siempre a menor frecuencia que los potenciales alelos parentales. Las frecuencias relativas de los alelos reales variaron entre quistes (como se detalla más adelante en el apartado referido a la dosis alélica). Sin embargo, ninguna secuencia identificada como quimera en un quiste fue detectada a una frecuencia mayoritaria en otro quiste.
- ii) El número de recombinantes detectados dependió del número de diferencias nucleotídicas existentes entre los potenciales alelos parentales. Tal y como se puede apreciar en la Tabla 4.9 se detectaron casi todos los recombinantes posibles entre cada par de variantes consecutivas. Esto solo es posible asumiendo una elevadísima tasa de recombinación, lo que contradice las escasas evidencias de recombinación en *Giardia*.
- iii) El número de recombinantes observado en algunos quistes superó ampliamente los cuatro alelos esperados como máximo en un único quiste, teniendo en cuenta que su dotación genética proviene de la replicación de dos núcleos diploides del trofozoito (ver Ciclo Vital).
- iv) Cuanto mayor fue la distancia física entre un par de variantes consecutivas, a mayor frecuencia se detectó la quimera resultante de la recombinación entre ellos.

Por último se detectaron niveles bajos de contaminación cruzada entre quistes. En algunos alineamientos se observaron alelos de otras muestras representados por un número muy reducido de lecturas. La mayoría de estas secuencias fueron eliminadas en el filtrado por frecuencia. Aquellas que alcanzaron el umbral necesario para ser consideradas alelos reales deben ser interpretadas con cautela, y se analizan y discuten en los apartos 4.2.4.2 y 5.4, respectivamente.

4.2.4.1.2 Confirmación por secuenciación Sanger

La autenticidad de las variantes detectadas por MPS se confirmó mediante la amplificación por separado de los siete loci, la clonación de los productos de PCR y la secuenciación Sanger de 16 clones por amplicón en el aislado M. En las amplificaciones se utilizó 1 μ L de la suspensión de quistes al igual que en el control positivo que se incluyó en la carrera de MPS.

Los resultados de la secuenciación masiva y la secuenciación Sanger mostraron una concordancia casi completa entre las dos metodologías (Tabla 4.12). Todos los alelos identificados en la secuenciación masiva del control M (> 100 quistes) se confirmaron por Sanger. Además de estos alelos, también se confirmó la existencia del alelo *rpL7a_6*, que fue detectado por MPS en seis quistes y el control positivo, pero solo pasó los filtros en el quiste M23 (Tabla 8.18, Anexos).

Tabla 4.12: Comparación de los resultados de MPS y secuenciación Sanger en el control del aislado M.

Alelos	Método de secuenciación			SANGER
	USEQ	AMPLIAN	AMPLISAS	
<i>bg_3</i>	354	755	1064	5
<i>bg_4</i>	308	641	545	2
<i>bg_5</i>	182	388	363	6
<i>bg_6</i>	174	353	331	1
<i>tpi_3</i>	135	271	417	9
<i>tpi_4</i>	69	199	263	7
<i>tpi_5</i>	—	—	—	—
<i>rpL7a_1</i>	79	155	271	6+3*
<i>rpL7a_5</i>	27	58	78	3+1*
<i>rpL7a_6</i>	—	6	—	2
<i>mlh1_5</i>	5	45	10	9
<i>mlh1_6</i>	17	53	27	6
<i>mlh1_7</i>	—	—	—	—
<i>mlh1_8</i>	2	—	4	1
<i>gdh_3</i>	15	78	57	5+q
<i>gdh_4</i>	—	42	—	10+q
<i>calt_4</i>	—	34	—	9
<i>calt_5</i>	—	8	—	3
<i>calt_6</i>	—	7	—	3
<i>fd_5</i>	—	16	—	10+1*
<i>fd_6</i>	—	8	—	5

Nota: para cada alelo se indican el número de lecturas asignadas por los distintos métodos de genotipado y el número de clones secuenciados por Sanger.

**singletons* respecto al alelo de referencia. *q*: quimera. Los guiones indican que el alelo no fue detectado. En negrita el alelo que no pasó los filtros de frecuencia en el control M (ver texto).

Otros dos alelos identificados por MPS en los quistes no se confirmaron por Sanger: *tpi_5* y *mlh1_7* (Tabla 8.18 y Tabla 4.12). El primero solo se había identificado en un quiste por lo que podría estar a tan baja frecuencia en el aislado que resultase indetectable por cualquiera de los dos métodos. El segundo se había identificado por MPS en los ocho quistes del aislado M pero no en el control. En su lugar se detectó el alelo *mlh1_8* cuya presencia fue confirmada por Sanger (Tabla 8.18 y Tabla 4.12). De hecho, este alelo *mlh1_8* también se había detectado en los quistes, pero había sido descartado como potencial recombinante entre *mlh1_6* y *mlh1_7*.

Por último, también se identificaron por Sanger siete *singletons*, cinco correspondientes con mutaciones que aparecieron en un único clon y dos recombinantes que no se detectaron ni en los quistes ni en el control por MPS, por lo que podría tratarse de variantes a baja frecuencia que se muestrearon al azar o artefactos generados durante el proceso de amplificación, clonación y secuenciación Sanger.

4.2.4.2 Genotipos

El análisis individual de siete loci en 40 quistes (8 x 5 aislados) produjo 277 genotipos monolocus en lugar de los 280 esperados debido a los tres casos en que la baja cobertura no permitió la identificación de alelos por ninguno de los tres métodos de genotipado: *fd* en D3 y *calt* en M22 y M24 (descritos previamente en el apartado 4.2.4.1).

En los quistes de un mismo aislado se detectaron los mismos alelos en casi todos los casos, contabilizándose tan solo 20 diferencias en los 277 amplicones genotipados. También fue frecuente la observación de un mismo alelo en distintos aislados, en contraposición a lo observado en los aislados del ensamblaje B en el estudio de diversidad (ver Experimento I), donde rara vez dos aislados compartieron haplotipos (Figura 4.13).

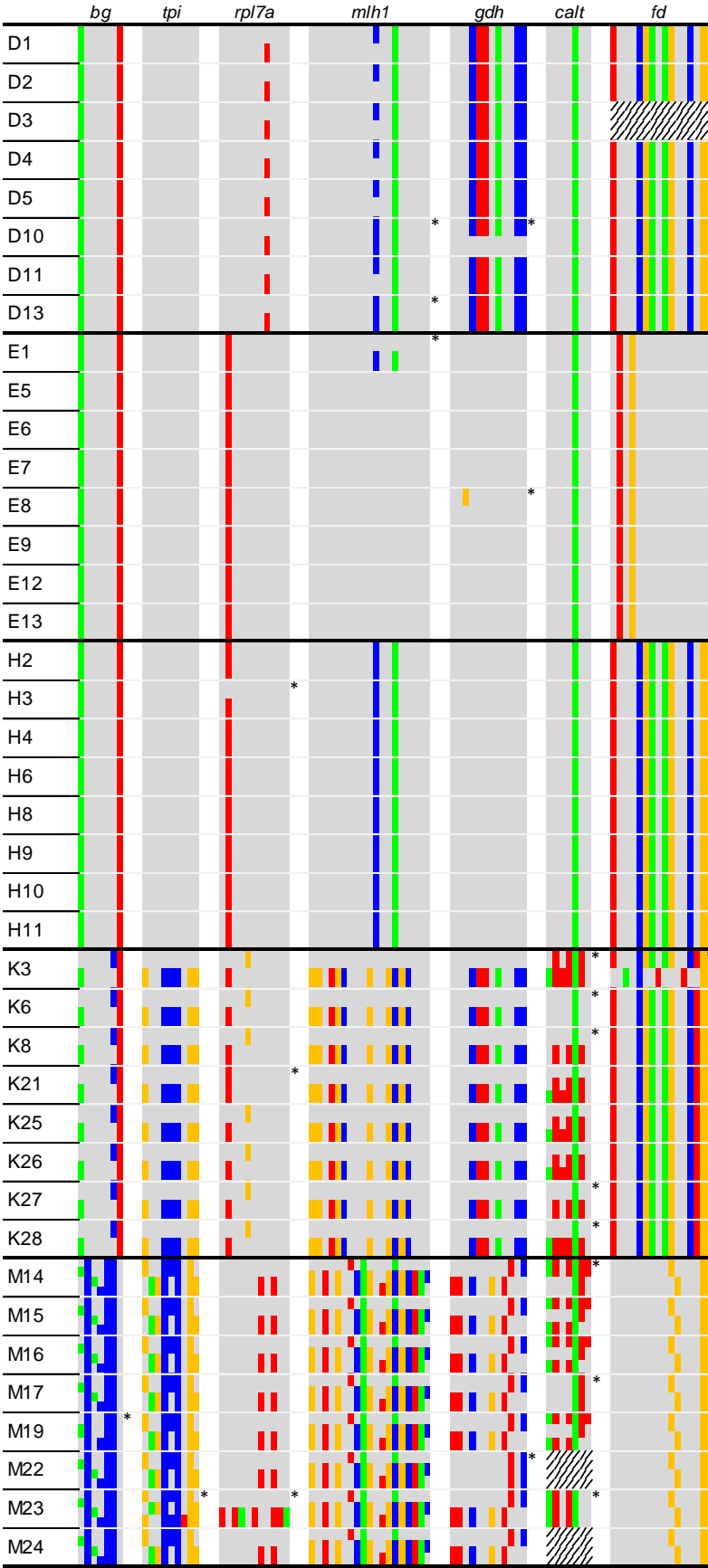


Figura 4.13: Genotipos de los quistes individuales.

Cada fila representa los genotipos de un quiste en los siete loci (indicados en la parte superior). Las columnas representan las posiciones variantes respecto al genoma de referencia (aislado GS). Para cada quiste se representan los haplotipos que pasaron los filtros de genotipado. Sin embargo, la fase en la que se encuentran los haplotipos de los distintos loci es desconocida, por lo que se establece una separación entre loci para no inducir a confusión.

Las posiciones conservadas respecto al genoma de referencia se representan en gris y las mutaciones siguiendo el siguiente código de colores: verde: adenina; azul: citosina; naranja: guanina; rojo: timina.

Las diferencias entre quistes de un mismo aislado aparecen marcadas con un asterisco.

Los casos en que no se produjeron secuencias que pasaran los filtros de genotipado en ninguno de los tres métodos aparecen señalados con un patrón rayado.

La diferencia más común entre quistes de un mismo aislado fue la pérdida de un alelo en un quiste heterocigoto: *bg_4* en M19, *gdh_4* en M22, *mlh1_2* en D10, *rpL7a_4* en K21, *mlh1_2* en D13 y *rpL7a_5* en M23. La inspección manual de los alineamientos de lecturas en el IGV, permitió detectar estos alelos en todos los casos, lo que sugiere que estaban presentes en los quistes. En los tres primeros casos, los alelos no se detectaron debido a que se descartaron durante el proceso de genotipado, ya que las variantes que los definen se encontraron a frecuencias claramente superiores al 7,5% establecido para que un alelo fuese considerado real. Por su parte, *rpL7a_4* y *mlh1_2* se encontraron a frecuencias cercanas al umbral en K21 y D13. La infrarrepresentación de estos alelos podría ser consecuencia de un efecto estocástico en la replicación durante los primeros ciclos de la PCR (*allele dropout*), por lo que fueron reconsiderados alelos reales. Esta misma explicación también podría justificar la pérdida de *rpL7a_5* en M23, aunque no se puede descartar que se trate de una diferencia real respecto al resto de quistes.

Se detectaron otros tantos casos de ganancia de alelos: *rpL7a_1* en H3, *gdh_2* en D10, *mlh1_1* en E1, *gdh_5* en E8, *fd_4* en K3 y *tpi_5* y *rpL7a_6* en M23. En los tres primeros casos estos alelos fueron mayoritarios en quistes de otros aislados, por lo que su presencia podría explicarse por una contaminación entre muestras durante la preparación de las librerías, una contaminación de los códigos de barras o un error en la lectura de los códigos de barras en la secuenciación. Por esta razón no se tuvieron en cuenta como diferencias reales entre quistes. El alelo *gdh_5* es un *singleton* que se diferencia del alelo mayoritario identificado en el propio quiste E8 en una única posición, por lo que se podría explicar por mutación o por una incorporación nucleotídica errónea durante la PCR.

En los tres últimos casos, se consideró que los nuevos alelos representaron diferencias reales respecto al resto de quistes del mismo aislado. El alelo *fd_4* se identificó únicamente en el quiste K3 y no hay indicios de que se encuentre a baja frecuencia en ningún otro quiste. Esto podría deberse a la pérdida de este alelo en los demás quistes por algún problema durante la PCR (*e.g.*, que exista alguna mutación en la región a la que se unen los cebadores que afecte a su capacidad de anillamiento) o a una contaminación desde alguno de los aislados que no se incluyeron en el análisis, pero no se descartó que se tratase de una diferencia real entre quistes.

El alelo *tpi_5* presenta una única mutación respecto a *tpi_2* (Tabla 8.8), por lo que podría haber sido considerado un *singleton* producido por un error de la polimerasa. Sin embargo, *tpi_2* no se identificó en el quiste M23, sino que es característico del aislado K. Por lo tanto no se pudo descartar que M23 sea un quiste ligeramente distinto al resto de quistes del aislado M.

Por último, el alelo *rpL7a_6* pasó los filtros únicamente en el quiste M23. La inspección manual de los alineamientos reveló la presencia de las variantes nucleotídicas características de este alelo en todos los quistes de este aislado a frecuencias comprendidas entre el 3% y el 8%. Además, este alelo se confirmó por Sanger en 2 de los 15 clones secuenciados, una frecuencia superior a la estimada tanto en un solo quiste (M23, AMPLIAN: 8.8%) como en el control de la muestra (AMPLIAN: 2.7%). Esto podría estar causado por una sobrerrepresentación aleatoria en la selección de los clones para la secuenciación Sanger, pero sugiere que debido a alguna característica propia del secuenciador, la secuenciación de este alelo funcionó peor en MPS, de modo que cabría la posibilidad de que este alelo se encuentre en realidad en todos los quistes del aislado M.

Un caso particular que merece un comentario aparte es el de *calt* en los quistes de los aislados K y M. En este caso las diferencias entre quistes se debieron a problemas en el genotipado derivados de la baja cobertura en este locus. Los quistes que presentaron mayor cobertura dentro de estos aislados mostraron los mismos tres alelos que los controles y que la secuenciación Sanger (Tablas 8.17, 8.18 y 4.12), mientras que en los que se obtuvieron menos lecturas se perdieron uno, dos o incluso los tres alelos. Por esta razón no se tuvieron en cuenta estos cambios como diferencias reales entre quistes.

4.2.4.3 Heterozigosidad

De los 277 genotipos inferidos, 116 fueron heterocigotos. Éstos no se distribuyeron aleatoriamente entre los individuos, como cabría esperar en una población panmíctica, sino que los quistes de los aislados E y H mostraron un genotipo homocigoto en todos los loci, y los quistes de los aislados K y M presentaron genotipos heterocigotos en todos los loci (con la excepción de *fd* en el aislado K)(Figura 4.13). Además, destacó la presencia de los alelos *bg_1*, *tpi_1*, *rpl7a_3*, *calt_1* y *gdh_2* en homocigosis en el aislado E, y en heterocigosis en el

aislado K. Estos dos resultados sugieren que la formación de genotipos homocigotos y heterocigotos podrían producirse en eventos que afectan al genoma completo y no a la segregación independiente de los distintos loci, como se discutirá más adelante.

Los quistes del aislado D mostraron un patrón intermedio, más próximo al de homocigosis de los aislados E y H, que al de heterocigosis de K y M. Presentaron un único alelo en *bg*, *tpi*, *gdh*, *calt* y *fd* y dos alelos en *rpl7a* y en *mlh1* que podrían haberse generado por mutación, ya que se diferenciaron entre sí en una única posición.

4.2.4.4 Dosis alélica

La dotación genética y el ciclo vital de *Giardia* limitan las posibles combinaciones de alelos en los quistes, ya que sus núcleos se forman por duplicación de los dos núcleos diploides de los trofozoitos. Por lo tanto, la dosis alélica se debe ajustar a unas proporciones teóricas. En el caso de los quistes que presentaron dos alelos, la relación entre los mismos podría ser 1:1 o 3:1 en favor de cualquiera de los dos alelos. En los quistes que presentaron tres alelos, la proporción sería 2:1:1 en favor de cualquiera de los tres alelos y en los que presentaron cuatro alelos el ciclo vital de *Giardia* predice una proporción 1:1:1:1.

El número de lecturas asignadas a cada alelo por cada método de genotipado debería ser proporcional al número de copias iniciales, lo que ofreció la posibilidad de inferir la dosis alélica de cada quiste a partir de la estimación de las frecuencias alélicas relativas (Figuras 8.1-8.7, anexos). La presencia de cuatro alelos de *bg* en los quistes del aislado M permitió establecer una comparación entre los valores observados y los valores esperados, ya que la única combinación teórica posible es 1:1:1:1, y por tanto la frecuencia esperada de cada alelo es 1/4. La comparación de las frecuencias absolutas obtenidas por los tres métodos de genotipado (número de lecturas) con los valores esperados reveló desviaciones significativas en todos los quistes (test χ^2 ; $p < 0.01$).

En la Figura 4.14 se representan las frecuencias relativas utilizadas en esta comparación. En ella destacan tres patrones principales: i) las frecuencias estimadas por los distintos métodos mostraron valores similares, lo que descarta que las diferencias respecto a las frecuencias esperadas se deban a un sesgo según el método de genotipado; ii) las frecuencias estimadas presentaron diferencias importantes respecto a la frecuencia esperada (0,25),

alcanzando valores próximos a 0,5 e inferiores a 0,1 (e.g., el alelo *bg_4*, ni siquiera pasó el filtro de frecuencia en M19) iii) las frecuencias relativas de los cuatro alelos mostraron una gran variación entre quistes, de forma que todos los alelos fueron mayoritarios en algún quiste, y minoritarios en otros, lo que descarta que las diferencias significativas se deban a un sesgo sistemático a favor de alguno de ellos, como la amplificación preferencial o la asociación de un error de secuenciación con alguna de las variantes.

Estos resultados cuestionaron la utilidad de las estimas de frecuencias alélicas ofrecidas por los distintos métodos de genotipado para predecir la dosis alélica de los quistes, lo que sugiere que se produjo un sesgo en la representación de los alelos durante la PCR, la construcción de las librerías y/o la secuenciación.

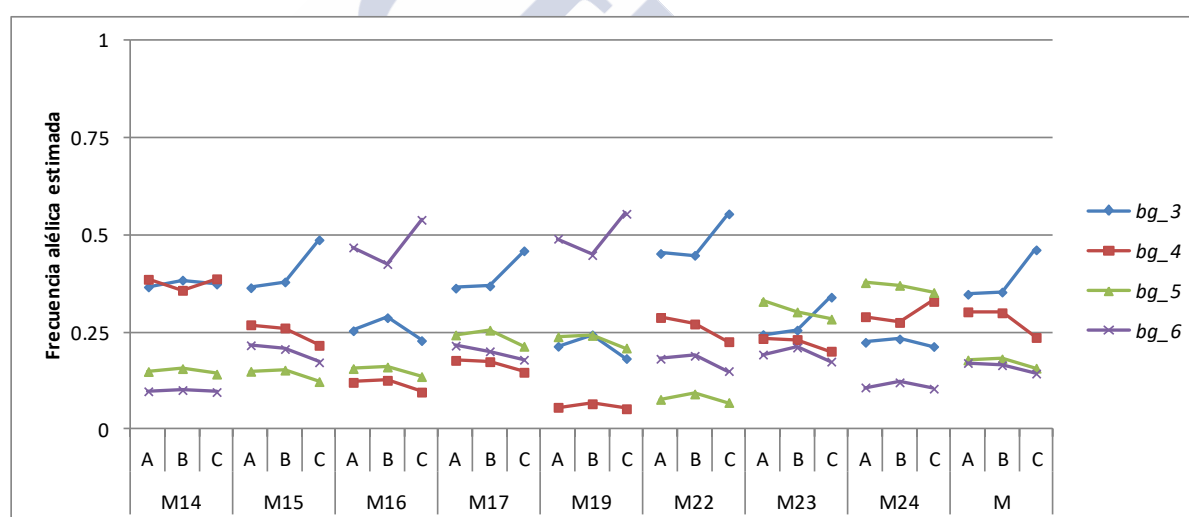


Figura 4.14: Estimas de la frecuencia alélica en el locus *bg* en quistes del aislado M. A: USEQ. B: AMPLIAN. C: AMPLISAS. Los colores identifican los distintos alelos tal y como se indica en la leyenda.

En los loci bialélicos y trialélicos, las estimas de frecuencia también ofrecieron un ajuste limitado respecto a las proporciones esperadas (Figuras 8.1-8.7). Por ejemplo, en el caso del aislado K, en el que todos los quistes presentaron dos alelos en *bg* (*bg_1/bg_2*), se observaron quistes con una proporción cercana a 1:1 (K25 y K27) y quistes próximos a 1:3 a favor de uno u otro alelo (K28: 0,27/0,73), lo que sugeriría que existen diferencias entre estos quistes. Sin embargo, también se encontraron quistes con frecuencias alélicas a medio camino entre dos

de los posibles escenarios (K6: 0.66/0.34 y K8: 0.35/0.65), que no permiten inferir una dosis alélica concreta.

En las Figuras 8.1-8.7 (Anexos) se observa cómo cada aislado se comportó prácticamente como un único bloque, donde los alelos detectados fueron siempre los mismos. Sin embargo, en los aislados que presentaron heterozigosidad, se observaron diferencias entre quistes. En algunos casos como en el locus *rpL7a* en el aislado D, fueron pequeñas variaciones entre quistes que podrían corresponderse con una misma dosis alélica, mientras que en otros casos, como en ese mismo locus en el aislado K, se observaron grandes variaciones que cubrieron todo el espectro de combinaciones posibles de dos alelos en un organismo tetraploide.

En función de estos resultados, se decidió no tener en cuenta las diferencias en las frecuencias alélicas al establecer comparaciones entre genotipos, ya que no se pudieron obtener estimas fiables de las dosis alélica.

4.2.4.5 Genotipos multilocus

La amplificación de los siete loci en quistes individuales permitió establecer de forma inequívoca la relación entre los genotipos obtenidos, permitiendo la inferencia de MLGs. Los MLGs descritos en este trabajo (Figura 4.13) se corresponden con el conjunto de alelos detectados en siete loci en un individuo sin tener en cuenta ni la dosis alélica (como se explica en el apartado anterior), ni la fase (la asociación física de los alelos de los distintos loci, ya sea en el mismo cromosoma, o en distintos cromosomas de un mismo núcleo).

Como resulta obvio no se observaron diferencias entre quistes de un mismo aislado más allá de las excepciones descritas previamente para cada locus en particular. En cambio, los quistes de cada aislado mostraron un MLG característico distinto al de los quistes de los demás aislados, lo que sugiere que el genotipado de estos siete loci podría ser una herramienta útil en la identificación de aislados.

Las combinaciones alélicas que se observaron en los distintos aislados ofrecieron señales claras de recombinación entre loci. Los quistes de los aislados E y H que fueron homocigotos para todos los loci aportaron la señal más evidente, ya que al no existir variación, necesariamente los alelos de los distintos loci se encontraron asociados en una misma

dotación haploide. Los loci *bg*, *tpi*, *rpl7a*, *calt* y *gdh* presentaron exactamente el mismo genotipo en los quistes de estas dos muestras, mientras que *mlh1* y *fd* presentaron genotipos distintos (Figura 4.13). Los alelos de *mlh1* en los dos aislados se diferenciaron en dos únicas posiciones, y no se podría descartar categóricamente que se pudiesen haber generado por mutación en lugar de por recombinación. Sin embargo, en el locus *fd*, presentaron dos alelos claramente distintos (10 sustituciones nucleotídicas), por lo que la presencia de estos dos alelos en combinación con el anterior bloque invariable solo se puede explicar mediante un evento de recombinación.

Los quistes de los aislados D y K compartieron un elevado número de alelos con los de los aislados E y H, de tal modo que el análisis de sus MLG proporcionó más evidencias de recombinación. Los quistes de los aislados D, E y H presentaron el mismo genotipo homocigoto en *bg*, *tpi*, y *calt*. Sin embargo, en *gdh* los quistes D presentaron el alelo *gdh_1* y los quistes E y H, el alelo *gdh_2*, con seis diferencias en su secuencia (Tabla 8.13, Anexos). Este hecho constituye en sí mismo una prueba de recombinación siguiendo la misma lógica utilizada en el párrafo anterior, pero además, los alelos de *bg*, *tpi*, y *calt*, se encontraron en heterocigosis en el aislado K, que sin embargo fue heterocigoto en *gdh* (*gdh_1/gdh_2*). Este resultado sugirió que no solo existe recombinación entre loci, sino que los alelos de un mismo locus se encuentran segregando en la población, ya que se observaron tanto en homocigosis como en heterocigosis.

Los quistes del aislado M presentaron un MLG muy particular, con entre dos y cuatro alelos por loci que no se observaron en ningún otro aislado con la excepción del alelo *rpl7a_1* que compartieron con los quistes del aislado D (Tablas 8.7-8.13). Del mismo modo que ocurrió en los casos anteriores y en base a las diferencias observadas entre los alelos de estos aislados en el resto de loci, la presencia de *rpl7a_1* en los quistes de ambos aislados solo se puede explicar a partir de un evento de recombinación.

4.2.4.6 Análisis filogenético

El análisis filogenético de las secuencias obtenidas reveló un patrón similar al observado en el ensamblaje B en el análisis de diversidad de los loci *gdh*, *bg* y *calt* (ver apartado 4.1.3.5). Todos los árboles generados mostraron una estructura poco definida con numerosos

grupos que presentaron valores bajos de *bootstrap*. De hecho, solo ocho grupos presentaron valores de *bootstrap* superiores al 75% (Figuras 8.8-8.14).

En varios loci apenas existieron secuencias de referencia disponibles con las que comparar los alelos detectados. En aquellos para los cuales se pudo incluir una parte representativa de la diversidad descrita hasta el momento, los alelos detectados se distribuyeron en distintas ramas del árbol (al igual que ocurría en el análisis de diversidad). Esto descartó la posibilidad de que la muestra analizada fuese representativa de una población aislada, sino que más bien al contrario se trató de una muestra diversa.

Tanto en las filogenias de *gdh* y *tpi* como en *mlh1* se observó una dicotomía inicial que separó a los distintos alelos en dos grupos principales. En el caso de *gdh* y *tpi* la comparación con las secuencias de referencia permitió identificar estos grupos con los subensamblajes BIII y BIV, aunque en ninguna de las dos filogenias mostraron valores de *bootstrap* superiores al 75%, lo que indica que el soporte estadístico para la existencia de los subensamblajes fue bajo. En las dos filogenias se observaron alelos de quistes heterocigotos en posiciones filogenéticamente alejadas, de modo que unos se asociaron con secuencias de referencia BIII y otros con secuencias BIV, a pesar de pertenecer a un mismo quiste. Esto ocurrió con los alelos *tpi_1* (BIV) y *tpi_2* (BIII), del aislado K (Figura 4.15) y los alelos *gdh_3* (BIV) y *gdh_4* (BIII), ambos detectados en los quistes M (Figura 4.16);. En el caso de *mlh1*, aunque no se encontraron disponibles secuencias de referencia para los subensamblajes BIII y BIV, se observó el mismo fenómeno. Los alelos *mlh1_3* y *mlh1_4*, presentes en los quistes K se encontraron alejados filogenéticamente, al igual que los alelos *mlh1_5* y *mlh1_6/mlh1_7* de los quistes del aislado M (Figura 4.17).

En todos los árboles se observó al menos un alelo (normalmente el que compartieron los aislados D, E, H y K) idéntico o muy similar al del genoma de referencia (GS) (Figuras 8.8-8.14). El otro genoma del ensamblaje B disponible, también se agrupó con el genoma de referencia, con la excepción de *fd*, donde *fd_2* (E) y *fd_5* (M) se agruparon con el genoma GS, mientras que *fd_3* (K) y *fd_1* (D y H) se agruparon con BAH15c1 formando el único grupo dentro del ensamblaje B con un *bootstrap* igual a 100. Las diferencias en la topología de los árboles filogenéticos en función del locus analizado se consideraron evidencias de la existencia de recombinación entre locus.

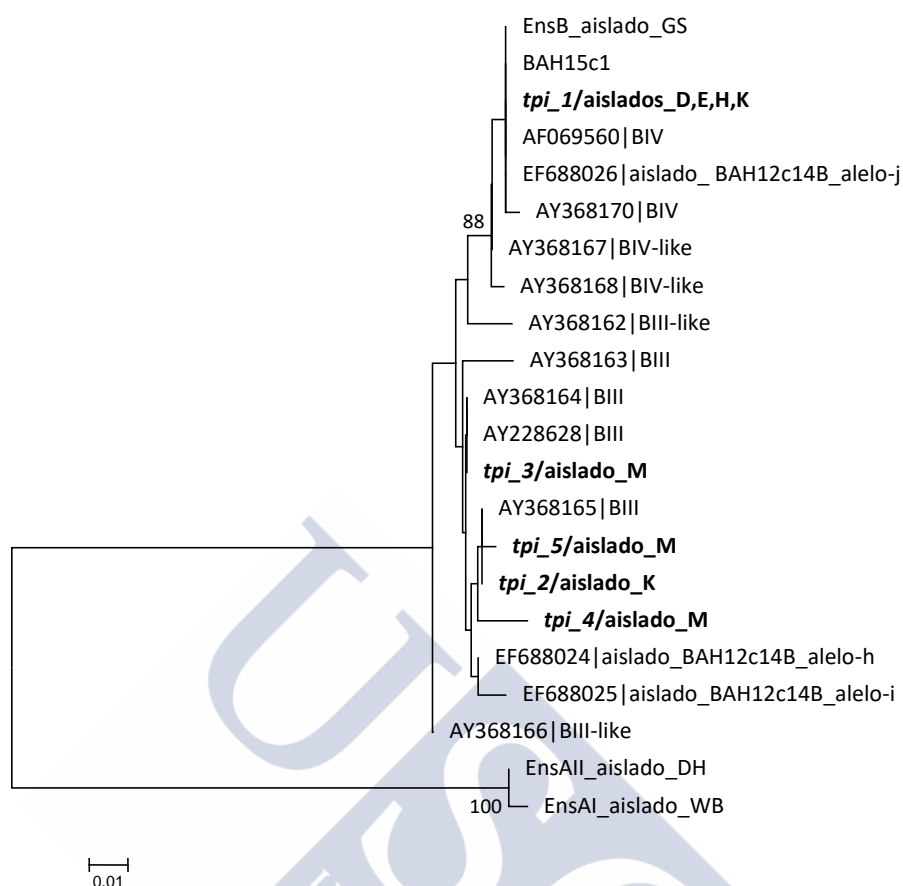


Figura 4.15: Árbol filogenético del locus *tpi*. Las relaciones filogenéticas fueron inferidas por el método NJ, y las distancias genéticas calculadas por el método de Tamura-Nei. Se utilizaron las secuencias de referencia de los aislados WB y DH del ensamblaje A como grupos externos. Se incluyeron como referencia las secuencias de los aislados GS BAH15c1 y BAH12c14B y las identificadas como variantes intragenotípicas por Wielinga y Thompson (2008). Los valores de *bootstrap* superiores al 75% se indican al lado de los nodos correspondientes. En negrita se indican los alelos identificados en este trabajo, así como el aislado en el que se detectaron. En los nombres de los taxones correspondientes a las secuencias de referencia se indica el número de acceso de GenBank.

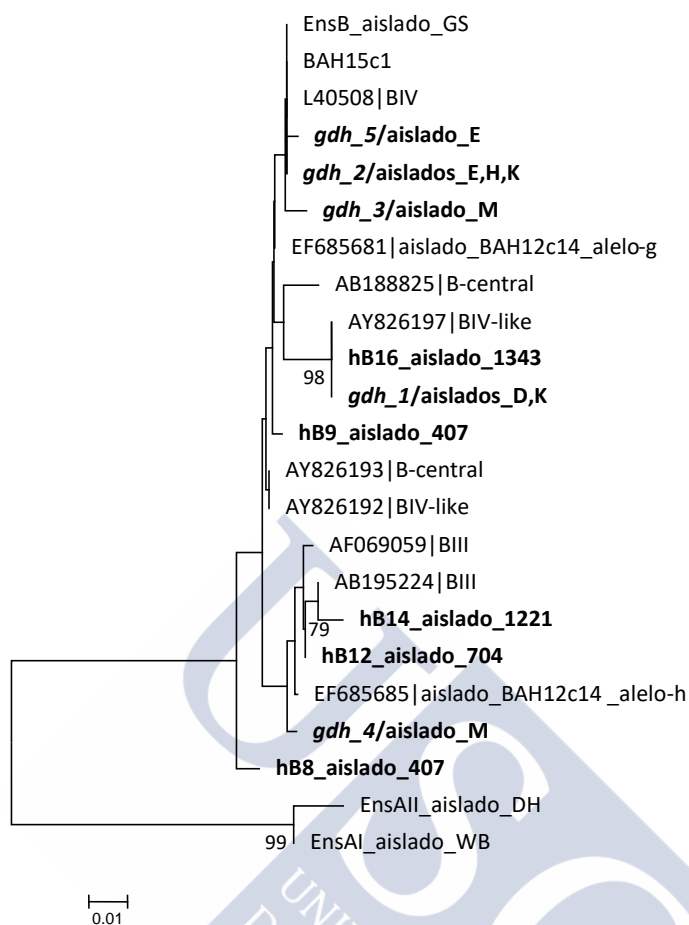


Figura 4.16: Árbol filogenético del locus *gdh*. Las relaciones filogenéticas fueron inferidas por el método NJ, y las distancias genéticas calculadas por el método de Tamura-Nei. Se utilizaron las secuencias de referencia de los aislados WB y DH del ensamblaje A como grupos externos. Se incluyeron como referencia las secuencias de los aislados GS, BAH15c1 y BAH12c14B y las identificadas como variantes intragenotípicas por Wielinga y Thompson (2008). Los valores de *bootstrap* superiores al 75% se indican al lado de los nodos correspondientes. En negrita se indican los alelos identificados en este trabajo, así como el aislado en el que se detectaron. En los nombres de los taxones correspondientes a las secuencias de referencia se indica el número de acceso de GenBank.

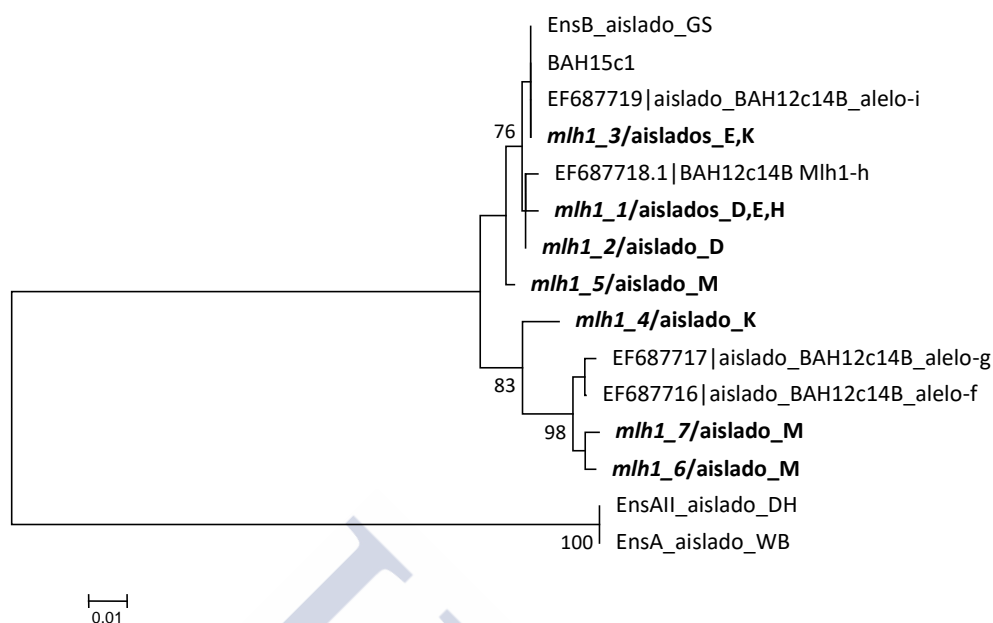


Figura 4.17: Árbol filogenético del locus *mlh1*. Las relaciones filogenéticas fueron inferidas por el método NJ, y las distancias genéticas calculadas por el método de Tamura-Nei. Se utilizaron las secuencias de referencia de los aislados WB y DH del ensamblaje A como grupos externos. Se incluyeron como referencia las secuencias de los aislados GS, BAH15c1 y BAH12c14B. Los valores de *bootstrap* superiores al 75% se indican al lado de los nodos correspondientes. En negrita se indican los alelos identificados en este trabajo, así como el aislado en el que se detectaron. En los nombres de los taxones correspondientes a las secuencias de referencia se indica el número de acceso de GenBank.





5. Discusión



5 Discusión

Lymbery y Thompson reconocen en su revisión “*The molecular epidemiology of parasite infections: tools and applications*” (Lymbery y Thompson, 2012) la siguiente reflexión: “Tibayrenc sugirió que los estudios de epidemiología molecular de las enfermedades parasitarias se pueden clasificar en dos tipos diferentes, dependiendo de si se ocupan únicamente de la identificación de los agentes causales de la enfermedad, o si consideran el impacto de la variabilidad genética en funciones posteriores (downstream functions), como las vías de transmisión, la infectividad, la virulencia o la resistencia a fármacos. Hasta la fecha, las técnicas moleculares han sido aplicadas principalmente para la identificación de los parásitos, en lugar de para el estudio de los patrones de transmisión y progresión de la enfermedad. Esto se debe en parte a la incapacidad de las técnicas moleculares para distinguir la variabilidad genética al nivel apropiado de resolución para tratar las funciones posteriores y en parte a la inadecuación de los métodos analíticos para interpretar la variación genética de un modo ecológicamente significativo”.

Este párrafo resume a la perfección la situación actual de la epidemiología molecular de la giardiasis y los problemas a los que se enfrenta. En un primer momento, la utilización de marcadores moleculares reveló la división de *G. duodenalis* en ocho ensamblajes, que probablemente se correspondían con distintas especies que hasta entonces habían permanecido ocultas bajo una misma identidad morfológica (Andrews *et al.*, 1989; Mayrhofer *et al.*, 1995; Monis *et al.*, 1999). Desde entonces, la atención se focalizó en la identificación de los ensamblajes y subensamblajes causantes de la infección (Cacciò *et al.*, 2002; Read *et al.*, 2002; Lalle *et al.*, 2005; Almeida *et al.*, 2006; Gelanew *et al.*, 2007; Robertson *et al.*, 2007; Lebbad *et al.*, 2008; Sahagún *et al.*, 2008). Mientras tanto, sigue existiendo un gran desconocimiento acerca de aspectos claves relativos a la epidemiología, la biología e incluso la patogenicidad del parásito, como su capacidad zoonótica (revisado en Feng y Xiao, 2011; Ryan y Cacciò, 2013), la existencia de recombinación y sus mecanismos (revisado en Cacciò y Sprong, 2010) o las causas de la sintomatología variable. Esto se debe en parte a la complejidad de la biología de este protozoo y en parte a las limitaciones de las técnicas moleculares y al uso inadecuado de las mismas.

Los análisis de alozimas y RFLPs, que se utilizaron para describir los ensamblajes y para su identificación en aislados de pacientes en las décadas de los 80 y los 90, carecen de la resolución necesaria para realizar una descripción precisa de la diversidad genética de las poblaciones de este parásito. El cambio a la secuenciación de regiones variables del genoma como técnica estándar para el genotipado de aislados resolvió parcialmente estos problemas de resolución. Sin embargo, esta técnica ha sido utilizada por lo general de un modo inapropiado para los fines pretendidos. La mayor parte de los estudios buscaron simplemente una secuencia que permitiese identificar la muestra, en ocasiones asociarla con un ensamblaje, en otras con un subensamblaje, y en los casos más ambiciosos con un genotipo concreto (o linaje), prácticamente como si se tratase de un cultivo bacteriano, en el que mediante una secuencia se pudiese identificar la “cepa” infectante. Pero *Giardia* no es un organismo haploide, ni sus infecciones son monoclonales, ni se conoce el papel que juega la recombinación en su ciclo vital, por lo que el genotipado mediante secuenciación directa de uno o varios loci tiene serias limitaciones de resolución (pérdida de variantes a baja frecuencia, incapacidad de establecer la combinación haplotípica) que impiden la correcta interpretación de los datos.

Para que los datos genéticos obtenidos de una muestra de *Giardia* sean útiles para el estudio de la genética de poblaciones –lo que Lymbery y Thompson (2012) denominan como “ecológicamente significativos”– deben permitir identificar tanto los alelos presentes en la población como sus frecuencias, e idealmente su asociación en genotipos dentro de cada individuo y las frecuencias genotípicas. La obtención de este tipo de datos es costosa y laboriosa pero necesaria, ya que la utilización de secuencias ambiguas de difícil interpretación lleva a un laberinto sin salida cuando se trata de confirmar la existencia de recombinantes o quistes heterocigotos en aislados en los que no se puede descartar una infección mixta (*e.g.*, Read *et al.*, 2004; Cacciò *et al.*, 2008; Sprong *et al.*, 2009; Lebbad *et al.*, 2011).

Numerosos autores han propuesto la secuenciación de quistes individuales como el camino a seguir para resolver estas dos cuestiones: el origen de la diversidad de las muestras y la existencia de aislados recombinantes:

- “El análisis de quistes únicos, aunque técnicamente complejo, tendría que ser realizado para establecer la contribución relativa de las infecciones mixtas y ASH a muestras heterogéneas.” (Cacciò *et al.*, 2008)

- “Nuevos métodos de tipado molecular, tales como la PCR de quistes individuales, son necesarios para investigar la direccionalidad y la frecuencia de la transmisión entre humanos y animales y para determinar la frecuencia de los eventos de recombinación dentro de los ensamblajes.” (Takumi *et al.*, 2012)
- “Es obvio que la distinción entre infecciones mixtas y verdaderos recombinantes es crucial. Esto requerirá del análisis de células únicas (i.e., quistes individuales) para ser formalmente indiscutible. Dado que el genotipado de quistes individuales es factible técnicamente (Miller and Sterling, 2007), y que hay disponibles ensayos de PCR específicos de ensamblaje (Geurden *et al.*, 2009; Almeida *et al.*, 2010), la investigación en esta dirección será de una importancia prioritaria.” (Thompson y Monis, 2011)
- “Idealmente, el genotipado debe realizarse a nivel de quistes individuales, ya que permitirá la diferenciación entre infecciones mixtas y la ocurrencia de recombinantes, porque existe la posibilidad de intercambio genético entre aislados del ensamblaje A e incluso entre aislados de los ensamblajes A y B.” (Feng y Xiao, 2011)

Durante el desarrollo de esta tesis se llevó a cabo un análisis cuantitativo de la diversidad genética en aislados de los ensamblajes A y B en el Área Sanitaria de Santiago de Compostela. A continuación, y a tenor de los resultados obtenidos, se diseñó e implementó un protocolo de genotipado multilocus de quistes individuales con objeto de identificar el origen de la variabilidad genética observada en los aislados del ensamblaje B, así como confirmar la existencia de mecanismos de intercambio genético entre quistes (recombinación).

Hasta donde sabemos, este es el primer análisis cuantitativo de la diversidad genética en aislados de *G. duodenalis* de pacientes humanos que se ha realizado en Galicia, y el primer estudio en el que se ha utilizado la clonación de productos de PCR en más de un locus en aislados de pacientes. Existen estudios previos similares en aislados adaptados al cultivo *in vitro* en laboratorio (Teodorovic *et al.*, 2007; Lasek-Nesselquist *et al.*, 2009), pero las condiciones de cultivo pueden sesgar los patrones de diversidad de las poblaciones (principalmente favorecer el crecimiento de unos linajes sobre otros), por lo que no es lo más adecuado para describir la diversidad y la estructura genética de las poblaciones naturales o de los aislados de cada paciente. También se han realizado estudios parecidos en aislados de pacientes, pero utilizando un único locus (Kosuwini *et al.*, 2010; Siripattanapipong *et al.*,

2011). La utilización de varios loci, permite obtener una visión más completa de los patrones de diversidad que afectan al genoma y hacer inferencias acerca de los efectos de la selección natural u otras fuerzas determinísticas sobre un locus concreto.

El análisis de quistes individuales es el más ambicioso de los realizados hasta la fecha, ya que incluye más loci y el uso de MPS permite la determinación de los haplotipos (Ankarklev *et al.*, 2012; Aguiar *et al.*, 2016).

5.1 DIVERSIDAD GENÉTICA EN LOS ENSAMBLAJES A Y B

Las estimas de diversidad genética mostraron grandes diferencias entre los ensamblajes A y B, sobre todo en las posiciones sinónimas. Estos lugares se utilizaron como una aproximación de las posiciones neutras del genoma, dado que las fuerzas evolutivas que actúan sobre ellas (tales como la selección en el uso de codones; *e.g.*, Maside *et al.*, 2004) tienen un efecto despreciable respecto a las que actúan sobre las posiciones no sinónimas (Charlesworth y Charlesworth, 2010). Además, en *Giardia* es especialmente difícil encontrar posiciones neutras ya que presenta un genoma muy compacto con escasas regiones no codificadoras y de pequeño tamaño, que de hecho podrían cumplir algún tipo de función reguladora, y por tanto estar también sometidas a selección (Morrison *et al.*, 2007; Adam *et al.*, 2013).

Las estimas de diversidad sinónima obtenidas concuerdan con las evidencias previas de una menor diversidad en el ensamblaje A en *gdh* y *bg* (Kosuwini *et al.*, 2010; Siripattanapipong *et al.*, 2011). Dado que los loci analizados no se encuentran ligados se puede inferir que nuestras observaciones son representativas del genoma de *Giardia*, por lo que las diferencias descritas sugieren que los ensamblajes A y B podrían haber experimentado una historia evolutiva reciente distinta.

Teniendo en cuenta que la diversidad neutra es directamente proporcional al tamaño de población y a la tasa de mutación (Kimura y Crow, 1964; Kimura, 1983), y asumiendo que ésta última es probablemente constante en los distintos ensamblajes y no varía de forma importante a lo largo del tiempo, los patrones de diversidad observados sugieren un tamaño efectivo de población menor en el ensamblaje A. Este hecho puede atribuirse a diferentes

propiedades ecológicas como una menor capacidad de proliferación en comparación con el ensamblaje B o a que su salto al hospedador humano ha sido más reciente.

En este sentido es interesante señalar que hay evidencias de que en el ensamblaje A existen alelos tan antiguos como los observados en el ensamblaje B, como demostró la descripción del subensamblaje AIII en animales salvajes (GenBank A.N.: EU637582 ;Cacciò *et al.*, 2008). Este subensamblaje mostró una distancia genética (d ; Tamura y Nei, 1993) en el locus *gdh* respecto a AI y AII superior a la máxima distancia observada entre haplotipos del ensamblaje B en este estudio (5,9% y 5,5%, respectivamente; vs. 4,2% entre hB1_aislado 321 y hB16_aislado1343). Sin embargo, apenas se ha detectado diversidad genética en aislados humanos más allá de las diferencias entre los subensamblajes AI y AII ($d = 1,02\%$). Esto apoyaría la hipótesis de que la población del *Giardia* del ensamblaje A que parasita humanos tiene un origen reciente y podría haberse originado por el salto de algún linaje desde un hospedador animal al hospedador humano.

Los casos de detección en humanos de los ensamblajes típicamente encontrados en animales de compañía o ganado, como C, D, E o F –como la detección del ensamblaje C en muestras clínicas en Brasil (Choy *et al.*, 2015) y de los ensamblajes C, D, E y F en distintos estudios revisados en (Feng y Xiao, 2011)— demostraron que existe una cierta plasticidad en la especificidad de hospedador que permite a los ensamblajes característicos de hospedadores animales, sobrevivir en el intestino humano. Estos eventos, aunque sean poco frecuentes, suponen una ventana de oportunidad para la adaptación de los distintos ensamblajes al ambiente intestinal de nuevos hospedadores y son una de las vías por las que el ensamblaje A podría haber dado el salto a la especie humana.

La posibilidad de que los ensamblajes fuesen capaces de adaptarse a nuevos hospedadores, tiene importantes implicaciones en el estudio de la transmisión del parásito. Si el salto a la especie humana se produjo después de la divergencia de los ensamblajes, no todos los linajes de los ensamblajes A y B habrían sido originalmente zoonóticos y cobraría especial importancia epidemiológica la identificación de dichos linajes, para el establecimiento de medidas de control de la salud pública. En este escenario hipotético, la existencia de recombinación dentro de los ensamblajes sería de enorme relevancia, ya que permitiría el intercambio de las características genéticas que permitieron la adaptación a un nuevo hospedador (en este caso humanos) entre linajes.

Los resultados en los test de Tajima no produjeron evidencias de que las poblaciones de los ensamblajes A o B se encontrasen alejadas del equilibrio mutación-deriva. Un estudio previo propuso una expansión poblacional reciente del ensamblaje A basándose en los valores negativos de la D de Tajima estimados a partir de secuencias del locus *tpi* sin distinguir posiciones sinónimas y no sinónimas (Choy *et al.*, 2015). Probablemente, este resultado se deba al efecto de la selección purificadora sobre las mutaciones no sinónimas, que las mantiene a baja frecuencia en las poblaciones, en lugar de a un evento poblacional, tal y como sugirieron los valores negativos de D en las posiciones no sinónimas en los datos obtenidos en esta tesis (significativamente distintos de cero en *gdh*-A, *bg*-A, *bg*-B y *calt*-B). En todo caso, los valores de D deben ser interpretados con cautela, ya que la detección de estructura en la población analizada, así como la posibilidad de que algunos de los *singletons* se correspondan con errores de la polimerasa durante la PCR, podrían haber afectado a la distribución de frecuencias de las variantes, y por tanto a las estimas de la D de Tajima.

En total, 86 de las 129 posiciones segregantes detectadas correspondieron a *singletons* (67%), siendo la proporción mayor en el ensamblaje A (85%). La utilización de protocolos de PCRs anidadas normalmente implica un aumento del número de ciclos, lo que incrementa la posibilidad de que se produzcan errores en la incorporación de nucleótidos durante la PCR. Sin embargo, el uso de una polimerasa de alta fidelidad con capacidad de corrección de errores y la observación de algunas de las mutaciones en secuencias de estudios previos (*e.g.*, *gdh*606C>T; *bg*198C>T, y *bg*496G>A), incluso a frecuencias polimórficas (*e.g.*, *bg*273A>G) (Kosuwin *et al.*, 2010; Siripattanapipong *et al.*, 2011; Ankarklev *et al.*, 2012), apoyan la autenticidad de las mismas.

Del mismo modo, es probable que otras variantes también hayan sido previamente detectadas en otros estudios, pero en la mayor parte de los casos la comparación con otros trabajos no es posible ya que el criterio utilizado para determinar qué variantes se deben considerar como verdaderas cuando se clonan los productos de PCR varía de unos estudios a otros. Por ejemplo, las nuevas variantes fueron consideradas verdaderas solo si: “ocurrieron dos o más veces en más de una librería de PCR o si aparecieron en todos los clones de una misma librería” (Lasek-Nesselquist *et al.*, 2009); si “fueron detectadas utilizando productos de PCR de dos amplificaciones independientes de la misma muestra de ADN” (Kosuwin *et al.*, 2010); o si “la sustitución en una posición ocurrió dos o más veces” (Siripattanapipong *et al.*,

2011). Sin embargo, en la secuenciación de un número limitado de clones, se espera que las mutaciones que existen en la población a baja frecuencia aparezcan representadas una única vez, y sea improbable detectarlas en dos “muestreos” independientes. Por esta razón, la exclusión de los *singletons* puede suponer también un sesgo de la D de Tajima, al provocar una infraestima de θ_w por la exclusión de las variantes a baja frecuencia. Por esta razón los *singletons* fueron incluidos en todos los análisis de diversidad genética y publicados en las bases de datos de secuencias para que estén a disposición de otros autores.

Otro factor que afecta a los valores de D , es la estructuración de la población: la mezcla de secuencias de poblaciones diferenciadas genéticamente provoca que las variantes que se encuentran fijadas en cada una de las poblaciones aparezcan a frecuencias intermedias en la muestra. Este fenómeno tiene un mayor efecto sobre π , cuyo valor se sobreestima en estos casos. Los análisis jerárquicos revelaron una estructuración significativa de la variación genética en ambos ensamblajes, de modo que las diferencias entre aislados contribuyeron a un 40% y un 60% de la varianza observada en los ensamblajes A y B, respectivamente. La existencia de diferencias fijadas entre aislados podría causar una sobreestima de la D de Tajima de los ensamblajes. Esto podría reflejarse en los valores positivos de D_s en el locus *gdh* del ensamblaje B, que es el que presentó mayor estructuración con un 85% de la varianza total causada por diferencias entre aislados, aunque no alcanzaron niveles significativos.

La diferenciación genética entre los distintos aislados también sugiere que cada infección está causada por una mínima fracción de la diversidad de la población del parásito. Además, si se tiene en cuenta que las muestras fueron recogidas en un área geográfica restringida (el Área Sanitaria de Santiago de Compostela ocupa unos 3400 km² e incluye alrededor de 440 000 habitantes), parece razonable pensar que la muestra analizada representa únicamente una fracción de la diversidad total de los ensamblajes A y B en Galicia. La comparación con las secuencias depositadas en las bases de datos mostró que los aislados analizados no forman una población diferenciada respecto a las poblaciones de otras localizaciones geográficas. En ausencia de datos poblacionales de otras partes del mundo que hubiesen permitido comparar las frecuencias alélicas en las poblaciones, y teniendo en cuenta que es probable que *solo* una parte de la diversidad haya sido recogida en nuestra muestra (como demuestran los nuevos haplotipos detectados en el análisis de quistes individuales), nos limitamos a comprobar si los haplotipos detectados habían sido secuenciados previamente en otras regiones. La presencia

en varios continentes de haplotipos idénticos o muy similares a los detectados en este estudio sugiere que no existe una diferenciación geográfica de la población de *G. duodenalis* que infecta humanos en el mundo.

5.2 DIVERSIDAD GENÉTICA DENTRO DE LOS AISLADOS

Giardia se consideró durante décadas un organismo de reproducción asexual, lo que potencialmente le conferiría la capacidad de acumular altos niveles de divergencia entre los alelos de los dos núcleos y entre los alelos de cada núcleo con el paso de las generaciones (Mark Welch y Meselson, 2000). Sin embargo, tanto en estudios previos (Morrison *et al.*, 2007; Franzén *et al.*, 2009; Lasek-Nesselquist *et al.*, 2009; Sprong *et al.*, 2009; Kosuwin *et al.*, 2010; Jerlström-Hultqvist *et al.*, 2010; Siripattanapipong *et al.*, 2011; Adam *et al.*, 2013; Huey *et al.*, 2013; Ankarklev *et al.*, 2015; de Lucio *et al.*, 2015; Minetti *et al.*, 2015) como en este trabajo se detectaron numerosos aislados monomórficos o casi monomórficos. Aislados como 122 y 147 del ensamblaje A o 704 y 1343 del ensamblaje B, no mostraron variación en ninguno de los tres loci más allá de algún que otro *singleton*. Este escenario sugiere la existencia de algún mecanismo de intercambio genético capaz de erosionar la heterozigosis esperada bajo reproducción estrictamente asexual, como podría ser la diplomixis (Poxleitner *et al.*, 2008).

Por el contrario otros aislados como 407 (B) y 1221 (B) mostraron niveles de diversidad elevados en varios loci, causada por la presencia de haplotipos muy distintos (tan divergentes como cualquier pareja de haplotipos extraídos de distintos aislados). En el ensamblaje A, en *bg* y *calt* se identificó un haplotipo común compartido por la mayoría de los aislados, y un haplotipo alternativo presente un número menor de aislados. La presencia de ambos haplotipos en algunos aislados podría indicar que se trataba de infecciones mixtas, provocadas por la coinfección de un mismo paciente por giardias procedentes de dos linajes monomórficos. Sin embargo, la comparación de los resultados de los distintos loci, reveló que los aislados en los que se detectó una mezcla de haplotipos no coincidieron (Tabla 5.1). Existen dos explicaciones alternativas para los patrones observados: la recombinación entre loci (que genera individuos con distintos genotipos en los distintos loci: aislados 209, 263 y 839) o la pérdida al azar durante la PCR (*allele dropout*) de uno de los haplotipos presentes en las infecciones mixtas (*bg*–hA1 en 209, *bg*–hA11 y *calt*–hA1 en 263 y *calt*–hA2 en 839) que

puede generar patrones genotípicos similares a los que produce la recombinación (Séré *et al.*, 2014).

Tabla 5.1: Combinaciones de haplotipos en los aislados del ensamblaje A.

Locus	Haplotipo	Aislados							
		122	147	152	209	251	263	321	839
<i>bg</i>	hA1	●	●	●		●	●	●	●
	hA11				○	○			○
<i>calt</i>	hA1	●	●	●	●	●		●	●
	hA2				○	○	○		

Nota: Los puntos negros representan la presencia en un aislado del alelo más común del ensamblaje A, mientras que los puntos blancos representan la presencia del alelo alternativo.

Este es el clásico escenario de inconsistencia entre los resultados de distintos loci que se observó habitualmente en estudios anteriores (Read *et al.*, 2004; Cacciò *et al.*, 2008; Sprong *et al.*, 2009). Diferenciar entre la pérdida de alelos en infecciones mixtas y la existencia de recombinantes es uno de los mayores problemas en el estudio de la variabilidad en aislados de *G. duodenalis* que exige el análisis de quistes individuales para ser resuelto (Cacciò *et al.*, 2008; Feng y Xiao, 2011; Thompson y Monis, 2011; Takumi *et al.*, 2012).

Las infecciones genotipadas como A+B mediante la amplificación específica del locus *tpi* se consideraron infecciones mixtas, ya que la existencia de híbridos A/B parece poco probable dada la elevada divergencia neutra observada entre los ensamblajes en todos los loci. La detección de dos secuencias quiméricas A+B en *calt* en el aislado 321, similares a las utilizadas previamente para proponer la existencia de recombinación entre ensamblajes (Teodorovic *et al.*, 2007; Lasek-Nesselquist *et al.*, 2009), debe ser interpretada con cautela. El hecho de que estos recombinantes fuesen secuenciados en un aislado donde las potenciales secuencias parentales se encontraban a mayor frecuencia sugiere que se trató de artefactos de PCR. La formación de quimeras en la amplificación mezclas de ADN fue estudiada previamente (Kanagawa, 2003) y confirmada aquí mediante el análisis MPS de quistes individuales. Su presencia en todos los quistes heterocigotos, unida al exceso de quimeras distintas detectadas en un único quiste (se llegaron a detectar hasta 10 quimeras, cuando el

número máximo de alelos es cuatro), demostró que se trata de un artefacto común en las amplificaciones por PCR.

5.3 AMPLIFICACIÓN Y SECUENCIACIÓN DE QUISTES INDIVIDUALES

El análisis molecular de células individuales en sistemas biológicos complejos es un área en expansión en los últimos años que ha logrado importantes avances en el conocimiento de la heterogeneidad de las células tumorales, la biología del desarrollo, la diversidad de los ecosistemas bacterianos (Kalisky y Quake, 2011; Blainey, 2013; Liang *et al.*, 2014; Wang y Navin, 2015), y jugará un papel principal en el estudio de la estructura poblacional y la diversidad genética de las infecciones parasitarias, como han mostrado estudios recientes en *Plasmodium* (Nair *et al.*, 2014), *Trypanosoma* (Valadares *et al.*, 2012) y también *Giardia* (Ankarklev *et al.*, 2012).

El éxito de la secuenciación de células individuales depende de dos procesos: el aislamiento celular y la amplificación del ADN. Tener la seguridad de que se amplificó el material genético de un único individuo resultó fundamental para poder extraer conclusiones precisas a partir de los resultados obtenidos. La presencia de más de una célula o de ADN libre de otros individuos en la PCR invalidaría cualquier resultado. En un primer momento se intentó aislar los quistes mediante citometría de flujo, pero no fue posible asegurar que en cada tubo de PCR se hubiese depositado un único quiste. La micromanipulación, a pesar de ser una técnica más laboriosa y lenta, permitió asegurar la separación de células individuales, mediante triple comprobación al microscopio, y establecer controles negativos precisos que descartaron la amplificación de ADN exógeno al quiste.

La amplificación del ADN en células individuales es el paso clave para obtener buenos resultados en la secuenciación y evitar la pérdida de alelos. El uso de un sistema de PCR directa supuso una gran ventaja respecto a métodos clásicos que incluyen un paso específico para la purificación de ADN, ya que minimizó la manipulación de los quistes y con ello las posibilidades de perder material genético y de contaminación con ADN exógeno (*e.g.*, humano, bacteriano, otras muestras,...). Es interesante destacar que, en nuestra experiencia, comparando los resultados de las amplificaciones realizadas durante los experimentos I y II,

la amplificación directa de quistes purificados ofreció un rendimiento claramente superior a la amplificación de ADN extraído de heces.

A pesar del éxito de la amplificación de quistes únicos, no se puede obviar la posibilidad de desequilibrios en la tasa de amplificación de los distintos loci. La optimización de la PCR multiplex en quistes individuales resultó muy compleja debido a la existencia de bandas de tamaño similar, a la necesidad de utilizar quistes aislados en el mismo día y al bajo número de copias de ADN molde de partida (que aumenta el efecto de las variaciones aleatorias en la tasa de replicación durante los primeros ciclos de la PCR). Por esta razón se decidió reamplificar el locus *tpi* que falló en la mayor parte de los quistes en lugar de intentar optimizar la reacción para conseguir su amplificación. Es importante destacar que *tpi* amplificó perfectamente tanto en los controles, como en los quistes del aislado K, lo que pone de manifiesto la variabilidad del rendimiento de las PCR multiplex de células únicas y las dificultades que supone su optimización. Finalmente, el resultado de la secuenciación reveló una tendencia a infrarrepresentar *calt* y *fd* en todos los quistes. Aun así, una de las principales ventajas de la secuenciación masiva en paralelo es el gran número de lecturas esperadas por locus y por muestra, lo que permitió detectar heterozigosidad alélica incluso en los loci más infrarrepresentados.

Las plataformas de MPS se basan en la secuenciación simultánea de millones de moléculas de ADN que han sido previamente clonadas sobre un soporte físico (en este caso las ISP). Entre las principales ventajas del uso de esta tecnología están: la rapidez del análisis, el alto rendimiento, la clonalidad de las lecturas y la escalabilidad de los experimentos (Buermans y Den Dunnen, 2014). Estas propiedades las convierten en idóneas para la secuenciación de amplicones en especies poliploides, ya que los haplotipos pueden ser definidos sin ambigüedad.

Mediante el uso de códigos de barras se incluyeron en una única reacción de secuenciación 40 quistes y 5 controles y se secuenciaron al mismo tiempo siete loci en cada uno de ellos. Un experimento del mismo tamaño, secuenciando ocho clones por locus, habría supuesto la secuenciación de más de 2520 colonias. En total, se obtuvieron más de 2 millones de lecturas clonales de las cuales apenas 190 000 eran idénticas a alguno de los alelos finalmente considerados verdaderos (Figura 4.12), lo que, aun así, supone una cobertura por locus inalcanzable para los métodos clásicos.

Por razones que ni nosotros ni los técnicos alcanzamos a explicar, la carga del chip mostró una eficiencia de aproximadamente el 50%, lo que supuso una pérdida de la mitad de las lecturas esperadas al usar un chip Ion-318v2. Aún así, la dimensión inicial del ensayo permitió obtener lecturas suficientes para determinar los genotipos de forma fiable prácticamente todos los amplicones.

5.4 DETERMINACIÓN DE LOS GENOTIPOS

La elevada concordancia entre los métodos de genotipado y la confirmación por Sanger de los alelos presentes en la muestra M es una prueba de la robustez de los genotipos obtenidos. A pesar de que USEQS y AMPLISAS tendieron a perder alelos cuando el número de lecturas largas fue reducido, el examen manual de los alineamientos de secuencias únicas anteriores al filtrado confirmó en la mayor parte de los casos la presencia de los alelos detectados por AMPLIAN, que fue el método más eficaz y preciso para la identificación de los mismos.

La baja cobertura de *calt* provocó problemas en la determinación de los genotipos, sobre todo en los quistes de los aislados K y M, en los que fue habitual la pérdida de alelos presentes en el control del aislado y en los quistes de mayor cobertura por falta de lecturas que cubriesen el amplicón al completo. En nuestra experiencia 200 lecturas completas fueron suficientes para que al menos AMPLIAN fuese capaz de determinar un genotipo correctamente. USEQ y AMPLISAS son métodos con una sensibilidad menor y precisan coberturas mayores para no perder alelos.

En los casos en los que la cobertura fue elevada, el método USEQ mostró mayores limitaciones que AMPLIAN y AMPLISAS, ya que fue más susceptible a los errores de secuenciación (como los *indels*). Esto pone de manifiesto que tanto AMPLIAN como AMPLISAS son herramientas más eficaces para la determinación de los genotipos, aunque cada uno de estos métodos tiene sus pros y sus contras: AMPLIAN exige un mayor tiempo de computación al tener que analizar cada locus en cada quiste por separado, pero a cambio es capaz de trabajar con niveles más bajos de cobertura, mientras que AMPLISAS es muchísimo más rápido pero requiere de un mayor número de lecturas largas. Sin duda, la comparación de los tres métodos y la revisión manual de los alineamientos aseguraron el máximo rigor en la

determinación de los genotipos. Además, la reproducibilidad de los resultados utilizando una metodología alternativa de clonación y secuenciación por Sanger demostró la veracidad de los genotipos obtenidos.

5.5 HETEROZIGOSIDAD vs. INFECCIONES MIXTAS

La principal fuente de variación dentro de los aislados fue la heterozigosidad de los quistes. Las aparentes diferencias entre quistes de un mismo aislado se produjeron en loci aislados y pudieron estar causadas por artefactos experimentales como *allele dropout*, errores en la asignación de los códigos de barras, o pequeños niveles de contaminación entre muestras.

En ningún caso se detectaron mezclas de quistes con genotipos multilocus claramente distintos dentro de un mismo aislado, por lo que se consideró que ninguno de ellos se correspondía con una infección mixta. Esto no descarta la existencia de infecciones mixtas, como se ha propuesto en estudios anteriores (e.g., Levecke *et al.*, 2009; Almeida *et al.*, 2010) y en este mismo trabajo en el caso de los aislados A+B. La confirmación de las infecciones mixtas y la estimación de su frecuencia dependerá de la aplicación de técnicas de genotipado de quistes individuales en un mayor número de aislados, ya que es el único mecanismo que permite distinguir entre la existencia de linajes híbridos o recombinantes y coinfecciones de distintos linajes de forma inequívoca.

La heterozigosidad observada en los quistes de los aislados K y M tiene dos posibles orígenes: la acumulación de mutaciones entre las distintas copias del genoma con el paso de las generaciones o el intercambio genético entre individuos distintos. Bajo la primera hipótesis, sería necesario buscar una justificación a la observación de quistes completamente homocigotos en los aislados E y H. Los alelos de los aislados E y H se observaron prácticamente en bloque en heterocigosis en el aislado K, lo que sugiere que, o bien las diferencias se acumularon entre núcleos y los genotipos homocigotos se originaron en una división reduccional (cada trofozoíto recibió dos copias del mismo núcleo), o bien los genotipos heterocigotos se formaron por la asociación de dos núcleos distintos en una misma célula, ya fuese por segregación de gametos o por fusión de dos trofozoítos (ver apartado siguiente).

Sin embargo estos procesos no han sido observados en ninguno de los análisis citológicos de la diferenciación y la división celular en *Giardia* que se han realizado hasta la fecha, que por el contrario han demostrado que la formación de los quistes se produce por replicación de los núcleos (y no por fusión de dos trofozoítos) y que tanto en la división mitótica de los trofozoítos como en la enquistación y la esquistación, los núcleos segregan de modo ecuacional (Sagolla *et al.*, 2006; Poxleitner *et al.*, 2008; Carpenter *et al.*, 2012; Jiráková *et al.*, 2012), lo que sugiere que podrían existir otros mecanismos de reproducción en *Giardia* que implican el intercambio de material genético entre individuos y que aún no han sido descritos.

Aunque estos resultados confirmaron la existencia de heterozigosidad en quistes individuales detectada anteriormente, contrastaron con lo observado por Ankarklev *et al.* (2012) en cuanto a las diferencias entre quistes. En dicho trabajo los quistes de un mismo aislado mostraron dobles picos en distintas posiciones. Este hecho fue interpretado como prueba de la existencia de quistes con distintos genotipos en un mismo aislado, aunque los propios autores reconocían que no podían descartar que se tratase de problemas para detectar las variantes en algunos quistes. Este tipo de diferencias entre quistes pudo ser un reflejo de los problemas de resolución que presenta la secuenciación directa, ya que es sencillo que algunas variantes pasasen inadvertidas (el alelo mayoritario enmascara al minoritario) y no se pudiesen identificar como un doble pico. También hay que tener en cuenta que el uso de dicha metodología impidió a los autores identificar los haplotipos presentes en los quistes.

En cualquier caso, no se puede descartar que existan diferencias en cuanto a la dosis alélica o la distribución de los alelos dentro de los distintos núcleos. El conocimiento de estos parámetros es importante para el estudio de la genética de poblaciones de este parásito y permitiría obtener una imagen más precisa del ciclo vital de *G. duodenalis*. Sin embargo, no se pudo obtener una estima fiable de las frecuencias alélicas, probablemente debido a los efectos estocásticos inherentes a la amplificación de un bajo número de copias, que pueden producir desbalance en las frecuencias alélicas durante la PCR, provocando incluso la pérdida de alelos. A pesar de ello, es importante destacar que los genotipos obtenidos en este estudio son mucho más precisos que los disponibles hasta la fecha y suponen un avance de cara al estudio de la asociación entre genotipos y fenotipos, así como en la comparación de los aislados de origen humano y animal para la determinación del potencial zoonótico de este parásito.

5.6 INTERCAMBIO GENÉTICO

En los últimos años ha habido un intenso debate acerca del modo de reproducción de los parásitos eucariotas (Heitman, 2010; Feretzaki y Heitman, 2013; Ramírez y Llewellyn, 2014; Tibayrenc y Ayala, 2015; Ramírez y Llewellyn, 2015; Weedall y Hall, 2015; Tibayrenc y Ayala, 2012). Aunque la reproducción sexual es una característica extendida a prácticamente todos los organismos eucariotas, en ausencia de evidencias citológicas de un ciclo sexual, muchos protozoos parásitos han sido considerados asexuales durante décadas. En algunos organismos –como los pertenecientes al filo Apicomplexa– se conoce desde hace tiempo una fase sexual (revisado en Smith *et al.*, 2002), mientras que las evidencias de intercambio genético han llevado a la identificación posterior de ciclos sexuales en otros como *Leishmania* o *Trypanosoma* (Heitman, 2010). En el caso de *G. duodenalis*, las evidencias de la existencia de un ciclo sexual son indirectas (Ramesh *et al.*, 2005; Cooper *et al.*, 2007; revisado en Birky, 2005; Logsdon, 2008; Birky, 2010; Andersson, 2012) y llevaron a la proposición de la existencia de un proceso parasexual en el que estarían implicados la fusión de los núcleos y el intercambio de material genético mediante recombinación mitótica o conversión génica: diplomixis (Poxleitner *et al.*, 2008).

La reproducción sexual, incluso a tasas muy bajas, presenta una serie de beneficios evolutivos como evitar la acumulación de mutaciones deletéreas en el genoma (Muller's Ratchet; Felsenstein, 1974; Muller, 1964) o favorecer la combinación en un mismo haplotipo de variantes ventajosas surgidas en distintos individuos. Aunque también tiene algunos costes como la ruptura de combinaciones alélicas exitosas, o la menor tasa reproductiva de la población, ya que normalmente se divide en dos sexos que tienen que contribuir para producir descendencia y son incapaces de hacerlo por separado (revisado en Butlin, 2002).

En los organismos que presentan reproducción sexual facultativa, la influencia del sexo en la evolución de sus poblaciones depende de la frecuencia y las condiciones en que éste se produce (Hartfield, 2016). Tibayrenc y Ayala han defendido desde 1990 la prevalencia de la reproducción clonal como norma general en la evolución de los protozoos parásitos, en lo que en un principio llamaron la Teoría clonal de los protozoos parásitos (Tibayrenc *et al.*, 1990) y más recientemente el modelo de Evolución predominantemente clonal (Tibayrenc y Ayala, 2012). Estos autores proponen que la recombinación se encuentra restringida por mecanismos genéticos endógenos y es tan escasa que no es capaz de subvertir el patrón prevalente de

clonalidad. En numerosos artículos (Tibayrenc y Ayala, 2002; Tibayrenc y Ayala, 2012; Tibayrenc y Ayala, 2014; Tibayrenc *et al.*, 1991) han remarcado que este modelo no se refiere al mecanismo citológico de reproducción sino que incide sobre las implicaciones que tienen la ausencia o la limitación de la recombinación sobre la estructura genética de las poblaciones: fuerte desequilibrio de ligamiento, mantenimiento espacio-temporal de los MLG y estructuración genética de las poblaciones en “*near-clades*” (grupos genéticos estables en simpatria, como podrían ser los ensamblajes). Este modelo incluye la homogamia y la autofecundación ya que según los autores sus consecuencias en la estructura genética de las poblaciones son similares a las de la reproducción asexual.

Por el contrario, algunos autores han argumentado que la aparente clonalidad observada podría ser producto de las limitaciones para detectar el intercambio genético en los protozoos patógenos como la baja resolución de los marcadores, la existencia de modelos de recombinación no mendeliana no detectables por los métodos analíticos clásicos o el estudio por separado de los distintos hospedadores con el riesgo de detectar únicamente MLG adaptados a cada hospedador que no reflejen la diversidad real de la especie patógena (Ramírez y Llewellyn, 2014).

En el análisis de los mecanismos de reproducción de *Giardia* a partir de nuestros resultados hay que distinguir entre dos patrones distintos: el que se observó al comparar los MLGs de los quistes dentro de cada aislado y el que se observó al comparar los MLGs de los distintos aislados.

Dentro de las infecciones se observaron los mismos MLGs en todos los quistes, lo que sugiere que durante la infección la reproducción es clonal. Los aislados homocigotos (E y H) no aportaron información al respecto, ya que en ausencia de variabilidad la reproducción sexual no genera genotipos distintos a los parentales. Pero los aislados K y M que mostraron heterocigosis en casi todos los loci, cumplieron una de las predicciones que derivan de la hipótesis de reproducción clonal: el exceso de heterocigotos.

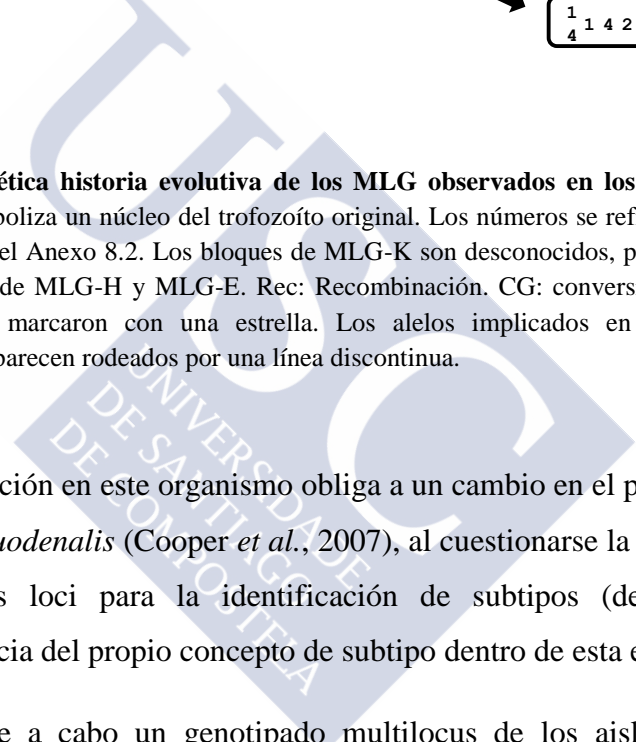
En un modelo de reproducción sexual bajo segregación mendeliana se espera que los distintos alelos presentes en una población se combinen de un modo aleatorio en cada generación, dando como resultado genotipos homocigotos y heterocigotos siguiendo las proporciones esperadas en el equilibrio Hardy-Weinberg. En un locus bialélico el escenario

más conservador ($p = q = 0,5$) predice un 12% de individuos homocigotos en organismos tetraploides. En los aislados que mostraron loci bialélicos (nueve casos) no se detectó ningún homocigoto (de 72 posibles), ya que en los cuatro homocigotos propuestos por los métodos de genotipado (Figura 4.13) la inspección manual de los alineamientos de lecturas en IGV reveló la presencia del segundo alelo.

La probabilidad de no muestrear una subpoblación de quistes a una frecuencia del 12% secuenciando ocho quistes por aislado es del 0,35; pero si tenemos en cuenta que este patrón se repitió en nueve casos independientes, la probabilidad de no encontrar homocigotos desciende a niveles significativos ($< 0,001$). Considerando que la clonalidad se definió como una restricción de la recombinación (Tibayrenc y Ayala, 2012) y la frecuencia de homocigotos se alejó notablemente de las expectativas en caso de panmixia, se pudo concluir que durante las infecciones predomina un modelo de reproducción clonal.

Al comparar los MLG de los quistes de distintos aislados se observaron evidencias claras de recombinación entre loci (Figura 4.13), en concordancia con los resultados obtenidos en estudios anteriores de *Giardia* (Cooper *et al.*, 2007). La diplomixis podría ser responsable de la ruptura de los MLG en los aislados E y H (iguales en cinco loci y distintos en *fd* y *mlh*), pero difícilmente puede explicar por sí sola que casi todos los alelos encontrados en estos aislados se mantengan en heterocigosis en los quistes del aislado K. Los datos apuntan a que o bien el MLG-K se formó por la combinación de un MLG similar al de E y H con otro MLG hipotético portador de los demás alelos, o bien los MLG de los quistes E y H se formaron en un evento de reducción de la diversidad por segregación conjunta de dos núcleos idénticos (padre-hijo) en contra de las evidencias citológicas (Carpenter *et al.*, 2012; Poxleitner *et al.*, 2008; Jiráková *et al.*, 2012).

En la Figura 5.1 se muestra un modelo hipotético en el cual el reordenamiento o la fusión de núcleos y la diplomixis pueden relacionar los MLGs observados en los aislados E, H y K en siete eventos biológicos sencillos. Para explicar los MLG completos es necesario inferir un MLG desconocido que explique el origen de alelos de E y H que no están presentes en K. El escenario que revelan estos datos de cara a la descripción de los patrones de evolución de los aislados del ensamblaje B se parece más a un puzzle al que faltan piezas, que a una filogenia dicotómica clásica. Es predecible que a medida que aumente el número de aislados secuenciados se irán encontrando los genotipos intermedios.



parecen rodeados por una línea discontinua.

La evidencia de recombinación en este organismo obliga a un cambio en el paradigma del genotipado de aislados de *G. duodenalis* (Cooper *et al.*, 2007), al cuestionarse la utilidad de la secuenciación de unos pocos loci para la identificación de subtipos (dentro de los ensamblajes) e incluso la vigencia del propio concepto de subtipo dentro de esta especie.

Por ejemplo, si se llevase a cabo un genotipado multilocus de los aislados E y H utilizando los loci más habituales *gdh*, *bg* y *tpi*, e incluyendo a mayores *rpL7a* y *calt* (más marcadores de los que se han utilizado en la inmensa mayoría de los estudios hasta la fecha), se llegaría a la conclusión de que ambos aislados son idénticos. Sin embargo, esta conclusión sería errónea ya que en *fd* y *mlhI* presentan alelos claramente distintos. Algo similar ocurriría si se comparasen estos aislados con el aislado D que presenta el mismo genotipo en *bg* y *tpi*, pero uno totalmente distinto en *gdh*.

En cambio, la existencia de recombinación abre la posibilidad de la identificación de los genes asociados con características epidemiológicas relevantes (factores de virulencia), ya que

aleatoriza el contexto genético en el que actúan los distintos alelos, permitiendo identificar de forma concreta el efecto que cada uno de ellos tiene por separado. Hasta el momento el genotipo de unos cuantos loci había sido interpretado como representativo del genoma completo y por tanto era indiferente identificar mutaciones en los genes causales o en cualquier otro, ya que se asumía que las variantes están ligadas y por tanto igualmente asociadas al fenotipo estudiado. La existencia de recombinación contradice este planteamiento y podría explicar la acumulación de resultados inconcluyentes en los intentos por asociar la sintomatología de la enfermedad con los genotipos clásicos (Aydin *et al.*, 2004; Gelanew *et al.*, 2007; Haque *et al.*, 2009; Homan y Mank, 2001; Minetti *et al.*, 2015; Mohammed Mahdy *et al.*, 2009; Puebla *et al.*, 2014; Read *et al.*, 2002; Sahagún *et al.*, 2008).

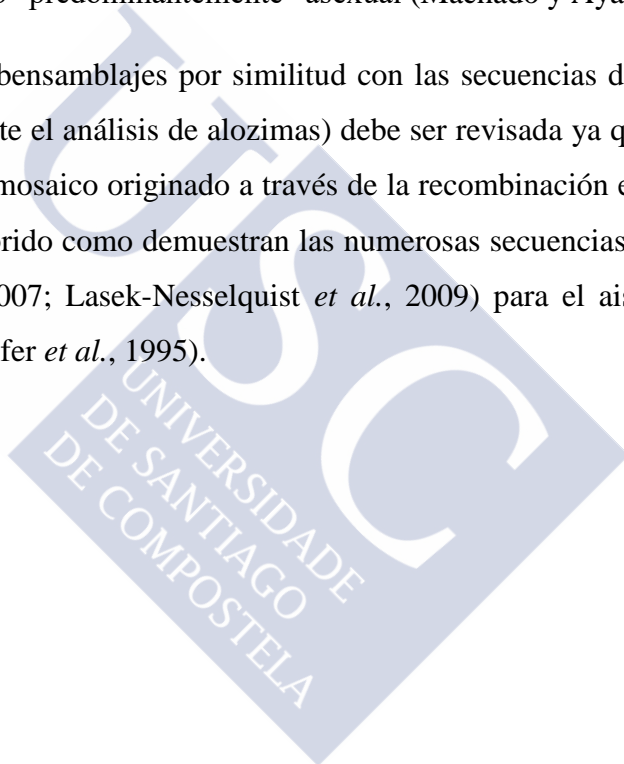
5.7 FILOGENIA

Las relaciones filogenéticas entre linajes se ven alteradas por la recombinación, lo que implica que las filogenias inferidas a partir de los distintos loci también pueden ser utilizadas para la detección de este tipo de eventos. La disposición de los alelos del aislado D en las distintas ramas de los árboles filogenéticos es un buen ejemplo de ello, así como la relación filogenética de los quistes E y H con el genoma de referencia según el locus que se analice (Figuras 8.8-8.14). Por otra parte, la existencia de recombinación limita la utilidad de los análisis filogenéticos y los árboles dicotómicos para inferir la historia evolutiva de las secuencias presentes en las poblaciones, como se ve reflejado en el bajo soporte estadístico de las ramas detectadas en las filogenias dentro de los ensamblajes A y B.

La división de las muestras del ensamblaje A en humanos en los subensamblajes AI y AII está ampliamente aceptada y así lo demuestran las filogenias obtenidas en el análisis de diversidad con la excepción de *bg* donde no existen diferencias entre estos subensamblajes en la región analizada. En el ensamblaje B la situación es mucho más compleja. Los numerosos grupos que se observaron en el análisis de diversidad se corresponden en su mayor parte con secuencias de un mismo aislado (Figuras 4.1, 4.2 y 4.3). Sin embargo, no se aprecia la división en subensamblajes BIII y BIV más que en el locus *gdh* donde los niveles de *bootstrap* muestran que esta división tiene un apoyo estadístico insuficiente (Figura 4.1).

Las filogenias inferidas a partir de los alelos detectados en los quistes individuales cuestionan la existencia de los subensamblajes BIII y BIV como unidades taxonómicas. La distribución de alelos de un mismo quiste en ramas opuestas de los árboles y agrupados con las secuencias de referencia de los subensamblajes BIII y BIV (Figuras 8.8-8.14) sugiere que existe tanto hibridación como recombinación entre ellos, ya que los alelos de un mismo quiste aparecieron asociados con las secuencias de referencia de uno u otro subensamblaje en distintos loci (*e.g.*, los alelos de los quistes K en *gdh* y *tpi*, Figuras 4.15 y 4.16). La existencia de hibridación entre linajes divergentes y recombinación entre linajes próximos ha sido previamente descrita en *Trypanosoma cruzi*, un protozoo parásito que también está considerado como un organismo “predominantemente” asexual (Machado y Ayala, 2001).

La identificación de los subensamblajes por similitud con las secuencias de los aislados de referencia (definidos mediante el análisis de alozimas) debe ser revisada ya que el genoma de estos aislados podría ser un mosaico originado a través de la recombinación entre distintos alelos, o incluso un genoma híbrido como demuestran las numerosas secuencias que han sido descritas (Teodorovic *et al.*, 2007; Lasek-Nesselquist *et al.*, 2009) para el aislado BIII de referencia: BAH12c14 (Mayrhofer *et al.*, 1995).





6. Conclusiones



6 Conclusiones

1. Las poblaciones de *G. duodenalis* de los ensamblajes A y B que parasitan humanos tienen una historia demográfica reciente distinta, y la primera presenta un tamaño efectivo de población menor.
2. Las infecciones causadas por *G. duodenalis* surgen a partir de un número reducido de individuos, que representan una pequeña fracción de la diversidad de dicha especie.
3. La secuenciación masiva en paralelo de quistes individuales permite la identificación precisa de genotipos multilocus.
4. Los aislados de *G. duodenalis* presentan un elevado nivel de diversidad genética, cuyo origen se debe a la heterozigosidad de los quistes (ASH) en mayor medida que a diferencias entre individuos (infecciones mixtas).
5. La recombinación genética es uno de los factores principales en la formación de nuevos genotipos.
6. Los subensamblajes BIII y BIV definidos en función de los loci *gdh* y *tpi*, no se corresponden con linajes independientes.





7. Bibliografía



7 Bibliografía

Orden SSI/445/2015 (2015) B.O.E.: 65: pages.

Adam, R. D. (2001). "Biology of *Giardia lamblia*." *Clin Microbiol Rev* **14**(3): 447-475.

Adam, R. D., E. W. Dahlstrom, C. A. Martens, D. P. Bruno, K. D. Barbian, S. M. Ricklefs, M. M. Hernandez, N. P. Narla, R. B. Patel, S. F. Porcella y T. E. Nash (2013). "Genome sequencing of *Giardia lamblia* genotypes A2 and B isolates (DH and GS) and comparative analysis with the genomes of Genotypes A1 and E (WB and pig)." *Genome Biol Evol* **5**(12): 2498-2511.

Aguiar, J. M., S. O. Silva, V. A. Santos, S. A. Taniwaki, T. M. Oliveira, H. L. Ferreira, L. B. Keid, F. Gregori y R. M. Soares (2016). "Evidence of heterozygosity and recombinant alleles in single cysts of *Giardia duodenalis*." *Rev Bras Parasitol Vet* **25**(2): 187-195.

Aguiar, J. M., S. O. d. Silva, V. A. n. d. Santos, G. A. R. Rossa, S. A. Taniwaki, F. Gregori, L. B. Keid y R. M. Soares (2015). "Multilocus amplification of genomic DNA from single cysts of *Giardia duodenalis* separated using micromanipulation technique." *Exp Parasitol* **157**: 84-87.

Almeida, A., E. Pozio y S. M. Cacciò (2010). "Genotyping of *Giardia duodenalis* cysts by new real-time PCR assays for detection of mixed infections in human samples." *Appl Environ Microbiol* **76**(6): 1895-1901.

Almeida, A. A., M. L. Delgado, S. C. Soares, A. O. Castro, M. J. Moreira, C. M. Mendonça, N. B. Canada y J. M. Correia Da Costa (2006). "Genotype analysis of *Giardia* isolated from asymptomatic children in northern Portugal." *J Eukaryot Microbiol* **53**(Suppl. 1): S177-178.

Amar, C. F. L., P. H. Dear y J. McLauchlin (2003). "Detection and genotyping by real-time PCR/RFLP analyses of *Giardia duodenalis* from human faeces." *J Med Microbiol* **52**(8): 681-683.

Andersson, J. O. (2012). "Double peaks reveal rare diplomonad sex." *Trends Parasitol* **28**(2): 46-52.

Andrews, R. H., M. Adams, P. F. Boreham, G. Mayrhofer y B. P. Meloni (1989). "*Giardia intestinalis*: electrophoretic evidence for a species complex." *Int J Parasitol* **19**(2): 183-190.

Ankarklev, J., O. Franzen, D. Peirasmaki, J. Jerlstrom-Hultqvist, M. Lebbad, J. Andersson, B. Andersson y S. G. Svard (2015). "Comparative genomic analyses of freshly isolated *Giardia intestinalis* assemblage A isolates." *BMC Genomics* **16**(1): 697-697.

Ankarklev, J., J. Jerlström-Hultqvist, E. Ringqvist, K. Troell y S. Svärd (2010). "Behind the smile: cell biology and disease mechanisms of *Giardia* species." *Nat Rev Microbiol* **8**(6): 413-422.

- Ankarklev, J., S. G. Svärd y M. Lebbad (2012). "Allelic sequence heterozygosity in single *Giardia* parasites." *BMC Microbiol* **12**(1): 65-65.
- Aurrecoechea, C., J. Brestelli, B. P. Brunk, J. M. Carlton, J. Dommer, S. Fischer, B. Gajria, X. Gao, A. Gingle, G. Grant, O. S. Harb, M. Heiges, F. Innamorato, J. Iodice, J. C. Kissinger, E. Kraemer, W. Li, J. A. Miller, H. G. Morrison, V. Nayak, C. Pennington, D. F. Pinney, D. S. Roos, C. Ross, C. J. Stoeckert, S. Sullivan, C. Treatman y H. Wang (2009). "GiardiaDB and TrichDB: Integrated genomic resources for the eukaryotic protist pathogens *Giardia lamblia* and *Trichomonas vaginalis*." *Nucleic Acids Res* **37**(Suppl. 1): D526-530.
- Aydin, A. F., B. A. Besirbellioglu, I. Y. Avci, M. Tanyuksel, E. Araz y A. Pahsa (2004). "Classification of *Giardia duodenalis* parasites in Turkey into groups a and B using restriction fragment length polymorphism." *Diagn Microbiol Infect Dis* **50**(2): 147-151.
- Bandelt, H. J., P. Forster y A. Röhl (1999). "Median-joining networks for inferring intraspecific phylogenies." *Mol Biol Evol* **16**(1): 37-48.
- Bartelt, L. a. y R. B. Sartor (2015). "Advances in understanding *Giardia*: determinants and mechanisms of chronic sequelae." *F1000prime reports* **7**(May): 62-62.
- Baruch, A. C., J. Isaac-Renton y R. D. Adam (1996). "The molecular epidemiology of *Giardia lamblia*: a sequence-based approach." *J Infect Dis* **174**(1): 233-236.
- Bernander, R., J. E. D. Palm y S. G. Svärd (2001). "Genome ploidy in different stages of the *Giardia lamblia* life cycle." *Cell Microbiol* **3**(1): 55-62.
- Birky, C. W. (2005). "Sex: Is *Giardia* doing it in the dark?" *Curr Biol* **15**(2): R56-58.
- Birky, C. W. (2010). "*Giardia* Sex? Yes, but how and how much?" *Trends Parasitol* **26**(2): 70-74.
- Blainey, P. C. (2013). "The future is now: Single-cell genomics of bacteria and archaea." *FEMS Microbiol Rev* **37**(3): 407-427.
- Bragg, L. M., G. Stone, M. K. Butler, P. Hugenholtz y G. W. Tyson (2013). "Shining a Light on Dark Sequencing: Characterising Errors in Ion Torrent PGM Data." *PLoS Comput Biol* **9**(4): e1003031-e1003031.
- Buermans, H. P. J. y J. T. Den Dunnen (2014). "Next generation sequencing technology: Advances and applications." *Biochim Biophys Acta* **1842**(10): 1932-1941.
- Butlin, R. (2002). "Evolution of sex: The costs and benefits of sex: new insights from old asexual lineages." *Nat Rev Genet* **3**(4): 311-317.
- Cacciò, S. M., R. Beck, M. Lalle, A. Marinculic y E. Pozio (2008). "Multilocus genotyping of *Giardia duodenalis* reveals striking differences between assemblages A and B." *Int J Parasitol* **38**(13): 1523-1531.

- Cacciò, S. M., M. De Giacomo y E. Pozio (2002). "Sequence analysis of the β -giardin gene and development of a polymerase chain reaction-restriction fragment length polymorphism assay to genotype *Giardia duodenalis* cysts from human faecal samples." *Int J Parasitol* **32**(8): 1023-1030.
- Cacciò, S. M. y U. Ryan (2008). "Molecular epidemiology of giardiasis." *Mol Biochem Parasitol* **160**(2): 75-80.
- Cacciò, S. M. y H. Sprong (2010). "*Giardia duodenalis*: Genetic recombination and its implications for taxonomy and molecular epidemiology." *Exp Parasitol* **124**(1): 107-112.
- Cano, L., A. de Lucio, B. Bailo, G. A. Cardona, A. S. Muadica, L. Lobo y D. Carmena (2016). "Identification and genotyping of *Giardia* spp. and *Cryptosporidium* spp. isolates in aquatic birds in the Salburua wetlands, Alava, Northern Spain." *Vet Parasitol* **221**: 144-148.
- Carpenter, M. L., Z. J. Assaf, S. Gourguechon y W. Z. Cande (2012). "Nuclear inheritance and genetic exchange without meiosis in the binucleate parasite *Giardia intestinalis*." *J Cell Sci* **125**(Pt 10): 2523-2532.
- Castro-Hermida, J. A., I. García-Presedo, A. Almeida, M. González-Warleta, J. M. Correia Da Costa y M. Mezo (2008). "Presence of *Cryptosporidium* spp. and *Giardia duodenalis* through drinking water." *Sci Total Environ* **405**(1-3): 45-53.
- Cavalier-Smith, T. (1987). "The origin of eukaryotic and archaebacterial cells." *Ann N Y Acad Sci* **503**: 17-54.
- Cavalier-Smith, T. (2003). "Protist phylogeny and the high-level classification of Protozoa." *Eur J Protistol* **39**(4): 338-348.
- Centro Nacional de Epidemiología, C. E. y. S. P. (2016). Resultados de la Vigilancia Epidemiológica de las enfermedades transmisibles. Informe Anual. Año 2014. Madrid: 89-91.
- Collier, S. A., L. J. Stockman, L. A. Hicks, L. E. Garrison, F. J. Zhou y M. J. Beach (2012). "Direct healthcare costs of selected diseases primarily or partially transmitted by water." *Epidemiol Infect* **140**(11): 2003-2013.
- Cooper, M. A., R. D. Adam, M. Worobey y C. R. Sterling (2007). "Population Genetics Provides Evidence for Recombination in *Giardia*." *Curr Biol* **17**(22): 1984-1988.
- Charlesworth, B. y D. Charlesworth (2010). Molecular Evolution and Variation. Elements of Evolutionary Genetics, Roberts and Company Publishers.
- Choy, S. H., M. A. K. Mahdy, H. M. Al-Mekhlafi, V. L. Low y J. Surin (2015). "Population expansion and gene flow in *Giardia duodenalis* as revealed by triosephosphate isomerase gene." *Parasit Vectors* **8**(1): 454-454.

- Dado, D., A. Montoya, M. A. Blanco, G. Miro, J. M. Saugar, B. Bailo y I. Fuentes (2012). "Prevalence and genotypes of *Giardia duodenalis* from dogs in Spain: possible zoonotic transmission and public health importance." *Parasitol Res* **111**(6): 2419-2422.
- de Lucio, A., R. Martinez-Ruiz, F. J. Merino, B. Bailo, M. Aguilera, I. Fuentes y D. Carmena (2015). "Molecular Genotyping of *Giardia duodenalis* Isolates from Symptomatic Individuals Attending Two Major Public Hospitals in Madrid, Spain." *PLoS One* **10**(12): 1-21.
- Durigan, M., A. G. a. Abreu, M. I. Zucchi, R. M. B. Franco y A. P. De Souza (2014). "Genetic diversity of *Giardia duodenalis*: Multilocus genotyping reveals zoonotic potential between clinical and environmental sources in a metropolitan region of Brazil." *PLoS One* **9**(12): e115489-e115489.
- ECDC, European Centre for Disease Prevention and Control (2017). "Surveillance Atlas of Infectious Diseases." Retrieved 22/01/2017, 2017, from <http://atlas.ecdc.europa.eu/public/index.aspx?Instance=GeneralAtlas>.
- Edgar, R. C. (2004). "MUSCLE: Multiple sequence alignment with high accuracy and high throughput." *Nucleic Acids Res* **32**(5): 1792-1797.
- Edlind, T. D. y P. R. Chakraborty (1987). "Unusual ribosomal RNA of the intestinal parasite *Giardia lamblia*." *Nucleic Acids Res* **15**(19): 7889-7901.
- Efstratiou, A., J. E. Ongerth y P. Karanis (2017). "Waterborne transmission of protozoan parasites: Review of worldwide outbreaks - An update 2011-2016." *Water Res* **114**: 14-22.
- Einarsson, E., S. Ma'ayeh y S. G. Svärd (2016). "An up-date on *Giardia* and giardiasis." *Curr Opin Microbiol* **34**: 47-52.
- Elias, E. V., R. Quiroga, N. Gottig, H. Nakanishi, T. E. Nash, A. Neiman y H. D. Lujan (2008). "Characterization of SNAREs determines the absence of a typical Golgi apparatus in the ancient eukaryote *Giardia lamblia*." *J Biol Chem* **283**(51): 35996-36010.
- Embley, T. M. y R. P. Hirt (1998). "Early branching eukaryotes?" *Curr Opin Genet Dev* **8**(6): 624-629.
- Excoffier, L. y H. E. L. Lischer (2010). "Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows." *Mol Ecol Resour* **10**(3): 564-567.
- Felsenstein, J. (1974). "The evolutionary advantage of recombination." *Genetics* **78**(2): 737-756.
- Felsenstein, J. (2009). "Confidence Limits on Phylogenies : An Approach Using the Bootstrap." *Evolution* **39**(4): 783-791.

- Feng, Y. y L. Xiao (2011). "Zoonotic potential and molecular epidemiology of *Giardia* species and giardiasis." *Clin Microbiol Rev* **24**(1): 110-140.
- Feretzaki, M. y J. Heitman (2013). "Unisexual Reproduction Drives Evolution of Eukaryotic Microbial Pathogens." *PLoS Pathog* **9**(10).
- Fernández-Álvarez, A., A. Martín-Alonso, N. Abreu-Acosta, C. Feliu, J. P. Hugot, B. Valladares y P. Foronda (2014). "Identification of a novel assemblage G subgenotype and a zoonotic assemblage B in rodent isolates of *Giardia duodenalis* in the Canary Islands, Spain." *Parasitology* **141**(1469-8161 (Electronic)): 206-215.
- Filice, F. P. (1952). *Studies on the cytology and life history of a Giardia from the laboratory rat*, University of California Press.
- Fletcher, S. M., D. Stark, J. Harkness y J. Ellis (2012). "Enteric Protozoa in the Developed World: a Public Health Perspective." *Clin Microbiol Rev* **25**(3): 420-449.
- Franzén, O., J. Jerlström-Hultqvist, E. Castro, E. Sherwood, J. Ankarklev, D. S. Reiner, D. Palm, J. O. Andersson, B. Andersson y S. G. Svärd (2009). "Draft genome sequencing of *Giardia intestinalis* assemblage B isolate GS: Is human giardiasis caused by two different species?" *PLoS Pathog* **5**(8): e1000560-e1000560.
- García-Presedo, I., S. Pedraza-Díaz, M. González-Warleta, M. Mezo, M. Gómez-Bautista, L. M. Ortega-Mora y J. A. Castro-Hermida (2013). "The first report of *Cryptosporidium bovis*, *C. ryanae* and *Giardia duodenalis* sub-assemblage A-II in roe deer (*Capreolus capreolus*) in Spain." *Vet Parasitol* **197**(3-4): 658-664.
- Gelanew, T., M. Lalle, A. Hailu, E. Pozio y S. M. Cacciò (2007). "Molecular characterization of human isolates of *Giardia duodenalis* from Ethiopia." *Acta Trop* **102**(2): 92-99.
- Gerwig, G. J., J. A. van Kuik, B. R. Leeftang, J. P. Kamerling, J. F. G. Vliegenthart, C. D. Karr y E. L. Jarroll (2002). "The *Giardia intestinalis* filamentous cyst wall contains a novel beta(1-3)-N-acetyl-D-galactosamine polymer: a structural and conformational study." *Glycobiology* **12**(8): 499-505.
- Geurden, T., E. Goossens, B. Levecke, F. Vercammen, J. Vercruysse y E. Claerebout (2009). "Occurrence and molecular characterization of *Cryptosporidium* and *Giardia* in captive wild ruminants in Belgium." *J Zoo Wildl Med* **40**(1): 126-130.
- Gil, H., L. Cano, A. de Lucio, B. Bailo, M. H. de Mingo, G. A. Cardona, J. A. Fernández-Basterra, J. Aramburu-Aguirre, N. López-Molina y D. Carmena (2017). "Detection and molecular diversity of *Giardia duodenalis* and *Cryptosporidium* spp. in sheltered dogs and cats in Northern Spain." *Infect Genet Evol* **50**: 62-69.
- Gómez-Muñoz, M. T., C. Cámara-Badenes, C. Martínez-Herrero Mdel, M. A. Dea-Ayuela, M. T. Pérez-Gracia, S. Fernández-Barredo, M. Santín y R. Fayer (2012). "Multilocus genotyping of *Giardia duodenalis* in lambs from Spain reveals a high heterogeneity." *Res Vet Sci* **93**(2): 836-842.

- Gómez-Muñoz, M. T., C. Navarro, M. M. Garijo-Toledo, M. A. Dea-Ayuela, S. Fernández-Barredo, M. T. Perez-Gracia, M. V. Domínguez-Márquez y R. Borrás (2009). "Occurrence and genotypes of *Giardia* isolated from lambs in Spain." *Parasitol Int* **58**(3): 297-299.
- Hall, T. (1999). "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT." *Nucleic Acids Symp Ser* **41**(41): 95-98.
- Hanevik, K., R. Bakken, H. R. Brattbakk, C. S. Saghaug y N. Langeland (2015). "Whole genome sequencing of clinical isolates of *Giardia lamblia*." *Clin Microbiol Infect* **21**(2): 192.e191-192.e193.
- Hanevik, K., K. A. Wensaas, G. Rortveit, G. E. Eide, K. Mørch y N. Langeland (2014). "Irritable bowel syndrome and chronic fatigue 6 years after giardia infection: a controlled prospective cohort study." *Clin Infect Dis* **59**(10): 1394-1400.
- Haque, R., D. Mondal, A. Karim, I. H. Molla, A. Rahim, A. S. Faruque, N. Ahmad, B. D. Kirkpatrick, E. Houpt, C. Snider y W. A. Petri Jr (2009). "Prospective case-control study of the association between common enteric protozoal parasites and diarrhea in Bangladesh." *Clin Infect Dis* **48**(9): 1191-1197.
- Haque, R., S. Roy, M. Kabir, S. E. Stroup, D. Mondal y E. R. Houpt (2005). "*Giardia* assemblage A infection and diarrhea in Bangladesh." *J Infect Dis* **192**(12): 2171-2173.
- Hartfield, M. (2016). "Evolutionary genetic consequences of facultative sex and outcrossing." *J Evol Biol* **29**(1): 5-22.
- Hashimoto, T., Y. Nakamura, T. Kamaishi, F. Nakamura, J. Adachi, K. Okamoto y M. Hasegawa (1995). "Phylogenetic place of mitochondrion-lacking protozoan, *Giardia lamblia*, inferred from amino acid sequences of elongation factor 2." *Mol Biol Evol* **12**(5): 782-793.
- Hashimoto, T., Y. Nakamura, F. Nakamura, T. Shirakura, J. Adachi, N. Goto, K. Okamoto y M. Hasegawa (1994). "Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place of a mitochondria-lacking protozoan, *Giardia lamblia*." *Mol Biol Evol* **11**(1): 65-71.
- Hashimoto, T., L. B. Sanchez, T. Shirakura, M. Muller y M. Hasegawa (1998). "Secondary absence of mitochondria in *Giardia lamblia* and *Trichomonas vaginalis* revealed by valyl-tRNA synthetase phylogeny." *Proc Natl Acad Sci U S A* **95**(12): 6860-6865.
- Heitman, J. (2010). "Evolution of eukaryotic microbial pathogens via covert sexual reproduction." *Cell Host and Microbe* **8**(1): 86-99.
- Heyworth, M. F. (2016). "*Giardia duodenalis* genetic assemblages and hosts." *Parasite* **23**: 13-13.
- Hill, W. G. y A. Robertson (1968). "Linkage disequilibrium in finite populations." *Theor Appl Genet* **38**(6): 226-231.

- Hlavsa, M. C., J. C. Watson y M. J. Beach (2005). "Giardiasis surveillance -- United States, 1998-2002." *MMWR Surveill Summ* **54**(3): 9-16.
- Homan, W. L., M. Gilsing, H. Bentala, L. Limper y F. Van Knapen (1998). "Characterization of *Giardia duodenalis* by polymerase-chain-reaction fingerprinting." *Parasitol Res* **84**(9): 707-714.
- Homan, W. L. y T. G. Mank (2001). "Human giardiasis: Genotype linked differences in clinical symptomatology." *Int J Parasitol* **31**(8): 822-826.
- Homan, W. L., F. H. J. van Enkevort, L. Limper, G. J. J. M. van Eys, G. J. Schoone, W. Kasprzak, A. C. Majewska y F. van Knapen (1992). "Comparison of *Giardia* isolates from different laboratories by isoenzyme analysis and recombinant DNA probes." *Parasitol Res* **78**(4): 316-323.
- Hörman, A., H. Korpela, J. Sutinen, H. Wedel y M. L. Hänninen (2004). "Meta-analysis in assessment of the prevalence and annual incidence of *Giardia* spp. and *Cryptosporidium* spp. infections in humans in the Nordic countries." *Int J Parasitol* **34**(12): 1337-1346.
- Horner, D. S. y T. M. Embley (2001). "Chaperonin 60 phylogeny provides further evidence for secondary loss of mitochondria among putative early-branching eukaryotes." *Mol Biol Evol* **18**(10): 1970-1975.
- Hudson, R. R. (2001). "Two-locus sampling distributions and their application." *Genetics* **159**(4): 1805-1817.
- Hudson, R. R. y N. L. Kaplan (1985). "Statistical properties of the number of recombination events in the history of a sample of DNA sequences." *Genetics* **111**(1): 147-164.
- Huey, C. S., M. A. K. Mahdy, H. M. Al-Mekhlafi, N. A. Nasr, Y. A. L. Lim, R. Mahmud y J. Surin (2013). "Multilocus genotyping of *Giardia duodenalis* in Malaysia." *Infect Genet Evol* **17**: 269-276.
- Hussein, A. I. A., T. Yamaguchi, K. Nakamoto, M. Iseki y M. Tokoro (2009). "Multiple-subgenotype infections of *Giardia intestinalis* detected in Palestinian clinical cases using a subcloning approach." *Parasitol Int* **58**(3): 258-262.
- Jerlström-Hultqvist, J., J. Ankarklev y S. G. Svärd (2010). "Is human giardiasis caused by two different *Giardia* species?" *Gut Microbes* **1**(6): 379-382.
- Jerlström-Hultqvist, J., O. Franzén, J. Ankarklev, F. Xu, E. Nohýnková, J. O. Andersson, S. G. Svärd y B. Andersson (2010). "Genome analysis and comparative genomics of a *Giardia intestinalis* assemblage E isolate." *BMC Genomics* **11**: 543-543.
- Jiráková, K., J. Kulda y E. Nohýnková (2012). "How Nuclei of *Giardia* Pass through Cell Differentiation: Semi-open Mitosis Followed by Nuclear Interconnection." *Protist* **163**(3): 465-479.

- Jukes, T. F. y C. R. Cantor (1969). Evolution of protein molecules. H. N. Munro. New York, Academic Press: 21-132.
- Kabnick, K. S. y D. A. Peattie (1990). "In situ analyses reveal that the two nuclei of *Giardia lamblia* are equivalent." *J Cell Sci* **95** (Pt 3): 353-360.
- Kalisky, T. y S. R. Quake (2011). "Single-cell genomics." *Nat Methods* **8**(4): 311-314.
- Kamikawa, R., Y. Inagaki, M. Tokoro, A. J. Roger y T. Hashimoto (2011). "Split introns in the genome of *Giardia intestinalis* are excised by spliceosome-mediated trans-splicing." *Curr Biol* **21**(4): 311-315.
- Kanagawa, T. (2003). "Bias and artifacts in multitemplate polymerase chain reactions (PCR)." *J Biosci Bioeng* **96**(4): 317-323.
- Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge Cambridgeshire ; New York, Cambridge University Press.
- Kimura, M. y J. F. Crow (1964). "The Number of Alleles That Can Be Maintained in a Finite Population." *Genetics* **49**: 725-738.
- Kosuwin, R., C. Putaporntip, U. Pattanawong y S. Jongwutiwes (2010). "Clonal diversity in *Giardia duodenalis* isolates from Thailand: Evidences for intragenic recombination and purifying selection at the beta giardin locus." *Gene* **449**(1-2): 1-8.
- Lal, A., S. Hales, N. French y M. G. Baker (2012). "Seasonality in human zoonotic enteric diseases: A systematic review." *PLoS One* **7**(4).
- Lalle, M., E. Pozio, G. Capelli, F. Bruschi, D. Crotti y S. M. Cacciò (2005). "Genetic heterogeneity at the β -giardin locus among human and animal isolates of *Giardia duodenalis* and identification of potentially zoonotic subgenotypes." *Int J Parasitol* **35**(2): 207-213.
- Lasek-Nesselquist, E., D. M. Welch y M. L. Sogin (2010). "The identification of a new *Giardia duodenalis* assemblage in marine vertebrates and a preliminary analysis of *G. duodenalis* population biology in marine systems." *Int J Parasitol* **40**(9): 1063-1074.
- Lasek-Nesselquist, E., D. M. Welch, R. C. A. Thompson, R. F. Steuart y M. L. Sogin (2009). "Genetic exchange within and between assemblages of *Giardia duodenalis*." *J Eukaryot Microbiol* **56**(6): 504-518.
- Lauwaet, T., B. J. Davids, D. S. Reiner y F. D. Gillin (2007). "Encystation of *Giardia lamblia*: a model for other parasites." *Curr Opin Microbiol* **10**(6): 554-559.
- Lebbad, M., J. Ankarklev, A. Tellez, B. Leiva, J. O. Andersson y S. Svärd (2008). "Dominance of *Giardia* assemblage B in León, Nicaragua." *Acta Trop* **106**(1): 44-53.
- Lebbad, M., I. Petersson, L. Karlsson, S. Botero-Kleiven, J. O. Andersson, B. Svenungsson y S. G. Svärd (2011). "Multilocus genotyping of human *Giardia* isolates suggests

- limited zoonotic transmission and association between assemblage B and flatulence in children." *PLoS Negl Trop Dis* **5**(8): e1262-e1262.
- Levecke, B., P. Geldhof, E. Claerebout, P. Dorny, F. Vercammen, S. M. Cacciò, J. Vercruysse y T. Geurden (2009). "Molecular characterisation of *Giardia duodenalis* in captive non-human primates reveals mixed assemblage A and B infections and novel polymorphisms." *Int J Parasitol* **39**(14): 1595-1601.
- Lewontin, R. C. (1964). "The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models." *Genetics* **49**(1): 49-67.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis y R. Durbin (2009). "The Sequence Alignment/Map format and SAMtools." *Bioinformatics* **25**(16): 2078-2079.
- Liang, J., W. Cai y Z. Sun (2014). "Single-Cell Sequencing Technologies: Current and Future." *J Genet Genomics* **41**(10): 513-528.
- Librado, P. y J. Rozas (2009). "DnaSP v5: A software for comprehensive analysis of DNA polymorphism data." *Bioinformatics* **25**(11): 1451-1452.
- Logsdon, J. M. (2008). "Evolutionary Genetics: Sex Happens in *Giardia*." *Curr Biol* **18**(2): R66-68.
- Lymbery, A. J. y R. C. A. Thompson (2012). "The molecular epidemiology of parasite infections: Tools and applications." *Mol Biochem Parasitol* **181**(2): 102-116.
- Lloyd, D. y J. C. Harris (2002). "*Giardia*: highly evolved parasite or early branching eukaryote?" *Trends Microbiol* **10**(3): 122-127.
- Machado, C. A. y F. J. Ayala (2001). "Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*." *Proc Natl Acad Sci U S A* **98**(13): 7396-7401.
- Mark Welch, D. y M. Meselson (2000). "Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange." *Science* **288**(5469): 1211-1215.
- Marti, M., A. Regös, Y. Li, E. M. Schraner, P. Wild, N. Müller, L. G. Knopf y A. B. Hehl (2003). "An Ancestral Secretory Apparatus in the Protozoan Parasite *Giardia intestinalis*." *J Biol Chem* **278**(27): 24837-24848.
- Maside, X., A. W. Lee y B. Charlesworth (2004). "Selection on Codon Usage in *Drosophila americana*." *Curr Biol* **14**(2): 150-154.
- Mateo, M., M. Mateo, A. Montoya, B. Bailo, J. M. Saugar, M. Aguilera, I. Fuentes y D. Carmena (2014). "Detection and molecular characterization of *Giardia duodenalis* in children attending day care centers in Majadahonda, Madrid, Central Spain." *Medicine* **93**(15): e75-e75.

- Mayrhofer, G., R. H. Andrews, P. L. Ey y N. B. Chilton (1995). "Division of *Giardia* isolates from humans into two genetically distinct assemblages by electrophoretic analysis of enzymes encoded at 27 loci and comparison with *Giardia muris*." *Parasitology* **111** (Pt 1: 11-17.
- McVean, G., P. Awadalla y P. Fearnhead (2002). "A coalescent-based method for detecting and estimating recombination from gene sequences." *Genetics* **160**(3): 1231-1241.
- Meloni, B. P., A. J. Lymbery y R. C. A. Thompson (1988). "Isoenzyme electrophoresis of 30 isolates of *Giardia* from humans and felines." *Am J Trop Med Hyg* **38**(1): 65-73.
- Mendonca, C., A. Almeida, A. Castro, M. de Lurdes Delgado, S. Soares, J. M. da Costa y N. Canada (2007). "Molecular characterization of *Cryptosporidium* and *Giardia* isolates from cattle from Portugal." *Vet Parasitol* **147**(1-2): 47-50.
- Miller, K. M. y C. R. Sterling (2007). "Sensitivity of nested PCR in the detection of low numbers of *Giardia lamblia* cysts." *Appl Environ Microbiol* **73**(18): 5949-5950.
- Minetti, C., K. Lamden, C. Durband, J. Cheesbrough, A. Fox y J. M. Wastling (2015). "Determination of *Giardia duodenalis* assemblages and multi-locus genotypes in patients with sporadic giardiasis from England." *Parasit Vectors* **8**(1): 444-444.
- Mohammed Mahdy, A. K., J. Surin, K. L. Wan, A. Mohd-Adnan, M. S. H. Al-Mekhlafi y Y. A. L. Lim (2009). "*Giardia intestinalis* genotypes: Risk factors and correlation with clinical symptoms." *Acta Trop* **112**(1): 67-70.
- Monis, P. T., R. H. Andrews, G. Mayrhofer y P. L. Ey (1999). "Molecular systematics of the parasitic protozoan *Giardia intestinalis*." *Mol Biol Evol* **16**(9): 1135-1144.
- Monis, P. T., R. H. Andrews, G. Mayrhofer y P. L. Ey (2003). "Genetic diversity within the morphological species *Giardia intestinalis* and its relationship to host origin." *Infect Genet Evol* **3**(1): 29-38.
- Monis, P. T., R. H. Andrews, G. Mayrhofer, J. Mackrill, J. Kulda, J. L. Isaac-Renton y P. L. Ey (1998). "Novel lineages of *Giardia intestinalis* identified by genetic analysis of organisms isolated from dogs in Australia." *Parasitology* **116** (Pt 1: 7-19.
- Monis, P. T., S. M. Caccio y R. C. A. Thompson (2009). "Variation in *Giardia*: towards a taxonomic revision of the genus." *Trends Parasitol* **25**(2): 93-100.
- Monis, P. T., G. Mayrhofer, R. H. Andrews, W. L. Homan, L. Limper y P. L. Ey (1996). "Molecular genetic analysis of *Giardia intestinalis* isolates at the glutamate dehydrogenase locus." *Parasitology* **112** (Pt 1: 1-12.
- Morrison, H. G., A. G. McArthur, F. D. Gillin, S. B. Aley, R. D. Adam, G. J. Olsen, A. a. Best, W. Z. Cande, F. Chen, M. J. Cipriano, B. J. Davids, S. C. Dawson, H. G. Elmendorf, A. B. Hehl, M. E. Holder, S. M. Huse, U. U. Kim, E. Lasek-Nesselquist, G. Manning, A. Nigam, J. E. J. Nixon, D. Palm, N. E. Passamaneck, A. Prabhu, C. I. Reich, D. S. Reiner, J. Samuelson, S. G. Svard y M. L. Sogin (2007). "Genomic

- minimalism in the early diverging intestinal parasite *Giardia lamblia*." *Science* **317**(5846): 1921-1926.
- Muller, H. J. (1964). "The relation of recombination to mutational advance." *Mutat Res* **1**: 2-9.
- Myers, S. R. y R. C. Griffiths (2003). "Bounds on the minimum number of recombination events in a sample history." *Genetics* **163**(1): 375-394.
- Naess, H., M. Nyland, T. Hausken, I. Follestad y H. I. Nyland (2012). "Chronic fatigue syndrome after *Giardia* enteritis: clinical characteristics, disability and long-term sickness absence." *BMC Gastroenterol* **12**(1): 13-13.
- Nair, S., S. C. Nkhoma, D. Serre, P. A. Zimmerman, K. Gorena, B. J. Daniel, F. Nosten, T. J. C. Anderson y I. H. Cheeseman (2014). "Single-cell genomics for dissection of complex malaria infections." *Genome Res* **24**(6): 1028-1038.
- Nash, T. E., D. A. Herrington, G. A. Losonsky y M. M. Levine (1987). "Experimental human infections with *Giardia lamblia*." *J Infect Dis* **156**(6): 974-984.
- Nash, T. E. y D. B. Keister (1985). "Differences in excretory-secretory products and surface antigens among 19 isolates of *Giardia*." *J Infect Dis* **152**(6): 1166-1171.
- Nash, T. E., T. McCutchan, D. Keister, J. B. Dame, J. D. Conrad y F. D. Gillin (1985). "Restriction-endonuclease analysis of DNA from 15 *Giardia* isolates obtained from humans and animals." *J Infect Dis* **152**(1): 64-73.
- Nei, M. (1987). *Molecular evolutionary genetics*. New York, Columbia University Press.
- Nei, M. y T. Gojobori (1986). "Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions." *Mol Biol Evol* **3**(5): 418-426.
- Nygård, K., B. Schimmer, Ø. Søbstad, A. Walde, I. Tveit, N. Langeland, T. Hausken y P. Aavitsland (2006). "A large community outbreak of waterborne giardiasis-delayed detection in a non-endemic urban area." *BMC Public Health* **6**: 141-141.
- Painter, J. E., J. W. Gargano, S. A. Collier, J. S. Yoder, C. Centers for Disease y Prevention (2015). "Giardiasis surveillance -- United States, 2011-2012." *MMWR suppl* **64**(3): 15-25.
- Philippe, H., P. Lopez, H. Brinkmann, K. Budin, A. Germot, J. Laurent, D. Moreira, M. Muller y H. Le Guyader (2000). "Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions." *Proc Biol Sci* **267**(1449): 1213-1221.
- Plutzer, J., J. Ongerth y P. Karanis (2010). "*Giardia* taxonomy, phylogeny and epidemiology: Facts and open questions." *Int J Hyg Environ Health* **213**(5): 321-333.
- Posada, D., K. A. Crandall y E. C. Holmes (2002). "Recombination in evolutionary genomics." *Annu Rev Genet* **36**: 75-97.

- Poxleitner, M. K., M. L. Carpenter, J. J. Mancuso, C.-J. R. C. J. R. Wang, S. C. Dawson y W. Z. Cande (2008). "Evidence for karyogamy and exchange of genetic material in the binucleate intestinal parasite *Giardia intestinalis*." *Science* **319**(5869): 1530-1533.
- Puebla, L. J., F. A. Núñez, Y. A. Fernández, J. Fraga, L. R. Rivero, I. A. Millán, L. A. Valdés y I. M. Silva (2014). "Correlation of *Giardia duodenalis* assemblages with clinical and epidemiological data in Cuban children." *Infect Genet Evol* **23**: 7-12.
- Quinlan, A. R. y I. M. Hall (2010). "BEDTools: A flexible suite of utilities for comparing genomic features." *Bioinformatics* **26**(6): 841-842.
- Ramesh, M. A., S. B. Malik y J. M. Logsdon (2005). "A phylogenomic inventory of meiotic genes: Evidence for sex in *Giardia* and an early eukaryotic origin of meiosis." *Curr Biol* **15**(2): 185-191.
- Ramírez, J. D. y M. S. Llewellyn (2014). "Reproductive clonality in protozoan pathogens-truth or artefact?" *Mol Ecol* **23**(17): 4195-4202.
- Ramírez, J. D. y M. S. Llewellyn (2015). "Response to Tibayrenc and Ayala: Reproductive clonality in protozoan pathogens - Truth or artefact?" *Mol Ecol* **24**(23): 5782-5784.
- Read, C., J. Walters, I. D. Robertson y R. C. A. Thompson (2002). "Correlation between genotype of *Giardia duodenalis* and diarrhoea." *Int J Parasitol* **32**: 229-231.
- Read, C. M., P. T. Monis y R. C. A. Thompson (2004). "Discrimination of all genotypes of *Giardia duodenalis* at the glutamate dehydrogenase locus using PCR-RFLP." *Infect Genet Evol* **4**(2): 125-130.
- Reboredo-Fernández, A., E. Ares-Mazás, S. M. Caccio y H. Gómez-Couso (2015). "Occurrence of *Giardia* and *Cryptosporidium* in wild birds in Galicia (Northwest Spain)." *Parasitology* **142**(7): 917-925.
- Reboredo-Fernández, A., E. Ares-Mazás, J. A. Martínez-Cedeira, R. Romero-Suances, S. M. Caccio y H. Gomez-Couso (2015). "*Giardia* and *Cryptosporidium* in cetaceans on the European Atlantic coast." *Parasitol Res* **114**(2): 693-698.
- Reboredo-Fernández, A., H. Gomez-Couso, J. A. Martínez-Cedeira, S. M. Caccio y E. Ares-Mazás (2014). "Detection and molecular characterization of *Giardia* and *Cryptosporidium* in common dolphins (*Delphinus delphis*) stranded along the Galician coast (Northwest Spain)." *Vet Parasitol* **202**(3-4): 132-137.
- Riley, D. E. y J. N. Krieger (1995). "Molecular and phylogenetic analysis of PCR-amplified cyclin-dependent kinase (CDK) family sequences from representatives of the earliest available lineages of eukaryotes." *J Mol Evol* **41**(4): 407-413.
- Robertson, L. J., T. Forberg, L. Hermansen, B. K. Gjerde y N. Langeland (2007). "Molecular characterisation of *Giardia* isolates from clinical infections following a waterborne outbreak." *J Infect* **55**(1): 79-88.

- Robertson, L. J., K. Hanevik, A. A. Escobedo, K. Mørch y N. Langeland (2010). "Giardiasis - why do the symptoms sometimes never stop?" *Trends Parasitol* **26**(2): 75-82.
- Robertson, L. J., L. Hermansen, B. K. Gjerde, E. Strand, N. Langeland y J. O. Alvsva (2006). "Application of Genotyping during an Extensive Outbreak of Waterborne Giardiasis in Bergen , Norway , during Autumn and Winter 2004 " *Appl Environ Microbiol* **72**(3): 2212-2217.
- Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz y J. P. Mesirov (2011). "Integrative genomics viewer." *Nat Biotechnol* **29**(1): 24-26.
- Rothberg, J. M., W. Hinz, T. M. Rearick, J. Schultz, W. Mileski, M. Davey, J. H. Leamon, K. Johnson, M. J. Milgrew, M. Edwards, J. Hoon, J. F. Simons, D. Marran, J. W. Myers, J. F. Davidson, A. Branting, J. R. Nobile, B. P. Puc, D. Light, T. a. Clark, M. Huber, J. T. Branciforte, I. B. Stoner, S. E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feierstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman, R. Kasinskas, T. Sokolsky, J. a. Fidanza, E. Namsaraev, K. J. McKernan, A. Williams, G. T. Roth y J. Bustillo (2011). "An integrated semiconductor device enabling non-optical genome sequencing." *Nature* **475**(7356): 348-352.
- Ryan, U. y S. M. Cacciò (2013). "Zoonotic potential of *Giardia*." *Int J Parasitol* **43**(12-13): 943-956.
- Sagolla, M. S., S. C. Dawson, J. J. Mancuso y W. Z. Cande (2006). "Three-dimensional analysis of mitosis and cytokinesis in the binucleate parasite *Giardia intestinalis*." *J Cell Sci* **119**(Pt 23): 4889-4900.
- Sahagún, J., A. Clavel, P. Goñi, C. Seral, M. T. Llorente, F. J. Castillo, S. Capilla, A. Arias y R. Gómez-Lus (2008). "Correlation between the presence of symptoms and the *Giardia duodenalis* genotype." *Eur J Clin Microbiol Infect Dis* **27**(1): 81-83.
- Saitou, N. y M. Nei (1987). "The neighbor-joining method: a new method for reconstructing phylogenetic trees." *Mol Biol Evol* **4**(4): 406-425.
- Schrader, C., A. Schielke, L. Ellerbroek y R. Johne (2012). "PCR inhibitors - occurrence, properties and removal." *J Appl Microbiol* **113**(5): 1014-1026.
- Sebastian, A., M. Herdegen, M. Migalska y J. Radwan (2016). "Amplisas: A web server for multilocus genotyping using next-generation amplicon sequencing data." *Mol Ecol Resour* **16**(2): 498-510.
- Séré, M., J. Kaboré, V. Jamonneau, A. M. G. Belem, F. J. Ayala y T. De Meeûs (2014). "Null allele, allelic dropouts or rare sex detection in clonal organisms: simulations and application to real data sets of pathogenic microbes." *Parasit Vectors* **7**(331): 1-13.
- Shen, H. E., L. Cao, J. Li, X. F. Tian, Z. H. Yang, Y. Wang, Y. N. Tian y S. Q. Lu (2011). "Visualization of chromosomes in the binucleate intestinal parasite *Giardia lamblia*." *Parasitol Res* **109**(5): 1439-1445.

- Siddall, M. E., H. Hong y S. S. Desser (1992). "Phylogenetic Analysis of the Diplomonadida (Wenyon, 1926) Brugerolle, 1975: Evidence for Heterochrony in Protozoa and Against *Giardia lamblia* as a "Missing Link"." *J Protozool* **39**(3): 361-367.
- Siripattanapipong, S., S. Leelayoova, M. Mungthin, R. C. A. Thompson, P. Boontanom, W. Saksirisampant y P. Tan-Ariya (2011). "Clonal diversity of the glutamate dehydrogenase gene in *Giardia duodenalis* from Thai isolates: evidence of genetic exchange or mixed infections?" *BMC Microbiol* **11**: 206-206.
- Smith, T. G., D. Walliker y L. C. Ranford-Cartwright (2002). "Sexual differentiation and sex determination in the Apicomplexa." *Trends Parasitol* **18**(7): 315-323.
- Sogin, M. L., J. H. Gunderson, H. J. Elwood, R. A. Alonso y D. A. Peattie (1989). "Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*." *Science* **243**(4887): 75-77.
- Sprong, H., S. M. Cacciò y J. W. B. Van Der Giessen (2009). "Identification of zoonotic genotypes of *Giardia duodenalis*." *PLoS Negl Trop Dis* **3**(12): e558-e558.
- Sulaiman, I. M., R. Fayer, C. Bern, R. H. Gilman, J. M. Trout, P. M. Schantz, P. Das, A. A. Lal y L. Xiao (2003). "Triosephosphate Isomerase Gene Characterization and Potential Zoonotic Transmission of *Giardia duodenalis*." *Emerg Infect Dis* **9**(11): 1444-1452.
- Tajima, F. (1989). "Statistical method for testing the neutral mutation hypothesis by DNA polymorphism." *Genetics* **123**(3): 585-595.
- Takumi, K., A. Swart, T. Mank, E. Lasek-Nesselquist, M. Lebbad, S. M. Cacciò y H. Sprong (2012). "Population-based analyses of *Giardia duodenalis* is consistent with the clonal assemblage structure." *Parasit Vectors* **5**(1): 168-168.
- Tamura, K. y M. Nei (1993). "Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees." *Mol Biol Evol* **10**(3): 512-526.
- Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei y S. Kumar (2011). "MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods." *Mol Biol Evol* **28**(10): 2731-2739.
- Teodorovic, S., J. M. Braverman y H. G. Elmendorf (2007). "Unusually low levels of genetic variation among *Giardia lamblia* isolates." *Eukaryot Cell* **6**(8): 1421-1430.
- Thompson, R. C. y P. Monis (2012). "*Giardia*-From Genome to Proteome." *Adv Parasitol* **78**: 57-95.
- Thompson, R. C. A. y A. Ash (2016). "Molecular epidemiology of *Giardia* and *Cryptosporidium* infections." *Infect Genet Evol* **40**: 315-323.
- Thompson, R. C. A. y P. T. Monis (2004). "Variation in *Giardia*: Implications for taxonomy and epidemiology." *Adv Parasitol* **58**: 69-137.

- Thompson, R. C. A. y P. T. Monis (2011). Taxonomy of *Giardia* Species. *Giardia A Model Organism*, Springer-Verlag Wien: 3-15.
- Tibayrenc, M. y F. J. Ayala (2002). "The clonal theory of parasitic protozoa: 12 years on." *Trends Parasitol* **18**(9): 405-410.
- Tibayrenc, M. y F. J. Ayala (2012). "Reproductive clonality of pathogens: a perspective on pathogenic viruses, bacteria, fungi, and parasitic protozoa." *Proc Natl Acad Sci U S A* **109**(48): E3305-3313.
- Tibayrenc, M. y F. J. Ayala (2014). "*Cryptosporidium*, *Giardia*, *Cryptococcus*, *Pneumocystis* Genetic Variability: Cryptic Biological Species or Clonal Near-Clades?" *PLoS Pathog* **10**(4): e1003908-e1003908.
- Tibayrenc, M. y F. J. Ayala (2015). "Reproductive clonality in protozoan pathogens - Truth or artifact? A comment on Ramírez and Llewellyn." *Mol Ecol* **24**(23): 5778-5781.
- Tibayrenc, M., F. Kjellberg, J. Arnaud, B. Oury, S. F. Brenière, M. L. Dardé y F. J. Ayala (1991). "Are eukaryotic microorganisms clonal or sexual? A population genetics vantage." *Proc Natl Acad Sci U S A* **88**(12): 5129-5133.
- Tibayrenc, M., F. Kjellberg y F. J. Ayala (1990). "A clonal theory of parasitic protozoa: the population structures of *Entamoeba*, *Giardia*, *Leishmania*, *Naegleria*, *Plasmodium*, *Trichomonas*, and *Trypanosoma* and their medical and taxonomical consequences." *Proc Natl Acad Sci U S A* **87**(7): 2414-2418.
- Tovar, J., G. León-Avila, L. B. Sánchez, R. Sutak, J. Tachezy, M. van der Giezen, M. Hernández, M. Müller y J. M. Lucocq (2003). "Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation." *Nature* **426**(6963): 172-176.
- Upcroft, J. A., K. G. Krauer y P. Upcroft (2010). "Chromosome sequence maps of the *Giardia lamblia* assemblage A isolate WB." *Trends Parasitol* **26**: 484-491.
- Valadares, H. M. S., J. R. Pimenta, M. Segatto, V. M. Veloso, M. L. Gomes, E. Chiari, K. J. Gollob, M. T. Bahia, M. de Lana, G. R. Franco, C. R. Machado, S. D. J. Pena y A. M. Macedo (2012). "Unequivocal identification of subpopulations in putative multiclonal *Trypanosoma cruzi* strains by FACs single cell sorting and genotyping." *PLoS Negl Trop Dis* **6**(7).
- van Keulen, H., R. R. Gutell, M. A. Gates, S. R. Campbell, S. L. Erlandsen, E. L. Jarroll, J. Kulda y E. A. Meyer (1993). "Unique phylogenetic position of Diplomonadida based on the complete small subunit ribosomal RNA sequence of *Giardia ardeae*, *G. muris*, *G. duodenalis* and *Hexamita* sp." *FASEB J* **7**(1): 223-231.
- Vickerman, K. (1982). Zoomastigophora. *Synopsis and Classification of Living Organisms*. S. P. Parker. New York, McGraw-Hill. **1**: 496-508.
- Wang, Y. y N. E. Navin (2015). "Advances and Applications of Single-Cell Sequencing Technologies." *Mol Cell* **58**(4): 598-609.

- Watterson, G. A. (1975). "On the number of segregating sites in genetical models without recombination." *Theor Popul Biol* **7**(2): 256-276.
- Weedall, G. D. y N. Hall (2015). "Sexual reproduction and genetic exchange in parasitic protists." *Parasitology* **142** Suppl(S1): S120-127.
- Wielinga, C., R. C. A. Thompson, P. Monis y U. Ryan (2015). "Identification of polymorphic genes for use in assemblage B genotyping assays through comparative genomics of multiple assemblage B *Giardia duodenalis* isolates." *Mol Biochem Parasitol* **201**(1): 1-4.
- Wielinga, C. M. y R. C. A. Thompson (2007). "Comparative evaluation of *Giardia duodenalis* sequence data." *Parasitology* **134**(Pt 12): 1795-1821.
- World Health Organization, W. (1996). *The world health report, 1996. Fighting disease, fostering development*. Geneva.
- World Health Organization, W. (2013). "Diarrhoeal Disease." Fact sheet nº330, 2017, from <http://www.who.int/mediacentre/factsheets/fs330/en/>.
- Xu, F., J. Jerlström-Hultqvist y J. O. Andersson (2012). "Genome-wide analyses of recombination suggest that *Giardia intestinalis* assemblages represent different species." *Mol Biol Evol* **29**(10): 2895-2898.
- Yoder, J. S. y M. J. Beach (2007). "Giardiasis surveillance--United States, 2003-2005." *MMWR Surveill Summ* **56**(7): 11-18.
- Yoder, J. S., J. W. Gargano, R. M. Wallace y M. J. Beach (2012). "Giardiasis surveillance United States, 2009-2010." *MMWR Surveill Summ* **61**(5): 13-23.
- Yoder, J. S., C. Haral y M. J. Beach (2010). "Giardiasis surveillance - United States, 2006-2008." *MMWR Surveill Summ* **59**(6): 15-25.
- Yong, T., K. Han, H. Yang y S. Park (2002). "PCR-RFLP analysis of *Giardia intestinalis* using a *Giardia*-specific gene, GLORF-C4." *Parasite* **9**(1): 65-70.
- Yu, L. Z., C. W. Birky y R. D. Adam (2002). "The two nuclei of *Giardia* each have complete copies of the genome and are partitioned equationally at cytokinesis." *Eukaryot Cell* **1**(2): 191-199.
- Zagordi, O., A. Bhattacharya, N. Eriksson y N. Beerenwinkel (2011). "ShoRAH: estimating the genetic diversity of a mixed sample from next-generation sequencing data." *BMC Bioinformatics* **12**(1): 119-119.



8. Anexos



8 Anexos

8.1 TABLAS DE HAPLOTIPOS DEL EXPERIMENTO I

Tabla 8.1: Haplotipos del locus *gdh* ensamblaje A.

	Aislados								Posiciones variantes																						
Haplotipos	122	147	152	209	251	263	321	839	276	323	335	385	390	401	404	419	469	470	479	529	532	539	563	603	621	660	679	691	699	753	
WB									C	T	T	G	T	G	A	T	G	C	A	T	G	A	C	T	C	C	A	A	T	C	
DH									C	T	C	T
hA1	8								C	T	T		
hA2		5		6	7		7	7	C	T	C	T	
hA3		6		8					-	-	-	C	T	C	-	
hA4			2			4			-	C	T	C	-	
hA5		1							.	.	C	C	T	C	T	
hA6		1							C	C	T	.	G	.	C	T		
hA7		1							-	-	-	A	C	T	C	-	
hA8		1							-	-	-	G	.	C	T	C	-	
hA9		1							T	.	C	T	C	T	
hA10			6						-	A	.	.	C	T	C	-	
hA11				1					T	C	C	T	C	T	
hA12				1					G	C	T	C	T	
hA13					1				A	.	.	C	T	C	T	
hA14						1			-	.	T	C	T	C	-	
hA15						1			-	T	C	T	C	-	
hA16						1			-	G	C	C	T	C	-	
hA17						1			-	C	T	T	.	G	C	-		
hA18							1		.	.	.	C	C	T	C	T	
hA19								1	A	C	T	C	T	

Nota: se utilizó como referencia el genoma de WB (ensamblaje A). Los puntos indican posiciones conservadas respecto al genoma de referencia. Los guiones indican huecos en el alineamiento. Para cada haplotipo se indican el número de clones por aislado. Las mutaciones no sinónimas se señalan en negrita.

Tabla 8.2: Haplotipos del locus *gdh* ensamblaje B.

Haplotipos	Aislados						Posiciones variantes																																	
	122	321	407	704	1221	1343	284	356	357	360	393	429	447	461	463	486	498	519	540	546	561	570	576	597	598	606	612	661	675	683	690	699	700	705	713	723	735	756	786	
GS							A	T	C	G	T	C	C	A	G	T	T	C	T	C	T	C	G	C	G	C	A	A	C	G	G	C	C	T	G	T	C	T	T	
hB1	2						-	.	T	.	.	T	T	.	A	.	.	T	C	T	C	G	T	.	.	.	C	.	-	-		
hB2	1						-	-	-	-	-	T	T	.	A	.	.	T	C	T	C	C	.	G	.	.	.	T	.	.	.	C	.	.	C	
hB3	6						-	-	-	-	-	T	T	.	A	.	.	T	C	T	C	G	T	.	.	.	C	.	.	C		
hB4	1						-	-	-	-	-	T	T	.	A	.	.	T	C	T	C	G	T	.	.	A	C	.	.	C		
hB5	1						-	-	-	-	-	T	T	.	A	.	.	T	C	T	C	T	G	<i>T</i>	.	.	T	.	.	C	.	.	C			
hB6	1						-	-	-	-	-	T	T	G	A	.	G	T	C	T	C	G	T	.	.	C	.	.	C			
hB7	1						-	-	-	-	-	T	T	G	A	.	.	T	C	T	C	G	T	.	.	C	.	.	C			
hB8	4						T	C	.	C	.	A	A	.	.	C	.	.	-			
hB9	2						C	T	-			
hB10	1						G	C	T	-			
hB11	1						C	T	T	.	-		
hB12			7				-	-	-	-	-	T	T	C	.	C	G	C	.	.	C			
hB13			1				-	-	-	-	-	T	T	C	.	C	G	.	.	A	C	.	.	C			
hB14				7			.	.	.	A	.	T	T	.	.	C	.	T	C	T	C	G	C	.	.	-			
hB15				1			.	.	.	A	C	T	T	.	.	C	.	T	C	T	C	G	C	.	.	-			
hB16					7		C	T	.	T	A	.	C	.	.	C	-				
hB17					1		.	C	C	T	.	T	A	.	C	.	.	C	-				

Nota: se utilizó como referencia el genoma de GS (ensamblaje B). Los puntos indican posiciones conservadas respecto al genoma de referencia. Los guiones indican huecos en el alineamiento. Para cada haplotipo se indican el número de clones por aislado. Las mutaciones no sinónimas se señalan en negrita y una mutación que introdujo un codón de parada se señala en negrita y cursiva.

Tabla 8.3: Haplotipos del locus *bg* ensamblaje A.

	Aislados								Posiciones variantes																														
Haplotipos	122	147	152	209	251	263	321	839	146	147	152	154	181	190	191	204	205	244	288	290	326	328	346	355	361	363	365	411	435	460	468	471	472	483	496	505	508	509	
WB/ DH									T	C	T	C	A	A	A	A	C	C	C	A	C	A	A	G	A	G	G	C	T	C	T	C	G	T	G	A	A	A	
*hA1	6	7	2		3	4	3	2
hA2	1								C
hA3	1								T
hA4		1							.	T
*hA5			1					
hA6				1					T
hA7					1				G
hA8						1			G	.	.
hA9							1			T
hA10								1	T
hA11				7	1			2	T	C
hA12					1				T	T	C
hA13						1			.	.	T	T	C
hA14							1		T	T	C
hA15								1	.	.	.	G
hA16									G	.	.
hA17					1			1	T
hA18						1			A	C
hA19							1		C
hA20								1	.	.	C	G
hA21									G
hA22									A	A
hA23									.	.	.	G
hA24									.	C
hA25									G
hA26								
hA27									C	.	.	.
hA28									A
hA29									A
hA30								1	G	G

Nota: se utilizó como referencia el genoma de WB (ensamblaje A). Los puntos indican posiciones conservadas respecto al genoma de referencia. Los guiones indican huecos en el alineamiento. Para cada haplotipo se indican el número de clones por aislado. Las mutaciones no sinónimas se señalan en negrita y una mutación que introdujo un codón de parada se señala en negrita y cursiva. * Haplotipos idénticos que se diferenciaron únicamente en la longitud de las secuencias.

Tabla 8.4: Haplotipos del locus *bg* ensamblaje B.

	Aislados						Posiciones variantes																										
Haplotipos	122	321	407	704	1221	1343	129	140	196	198	202	209	210	214	228	242	273	296	314	354	369	388	406	423	427	438	444	478	496	500	518	526	533
GS							T	A	G	C	G	T	T	A	A	G	A	T	A	T	T	A	A	C	A	C	G	G	G	A	A	A	A
hB1	5						-	-	-	-	-	-	-	-	C	C
hB2	1						-	-	-	-	-	-	-	-	.	A	.	.	C	C
hB3		3					C	C	C
hB4		2					C	C	C	C	.	T
*hB5		1					C	C	C	C
hB6		1					C	C	C
hB7		1					C	G	C
hB8		3					C	C
hB9		1					C	.	A	C	A
hB10		1					C	C	A
hB11		1					C	C	G
**hB12		3					C	C	.	G	C
hB13		1					C	C	.	G	C	G	.	.	.
**hB14		1					C	C	.	G	C
hB15		1					-	-	.	.	.	C	.	G	C
hB16		2					C	C	C
hB17		1					C	.	T	.	.	C	G	C
hB18		1					C	C	C	C	G	.
*hB19		1					C	C	C	C
hB20		1					C	A	C	C
hB21		1					C	.	.	.	A	C	A
hB22		3					C	C	T
hB23		1					C	C	T	.	A
hB24		1					C	C	.	G	.	T
hB25		1					C	G	G	.	.	C	.	.	.	C	T
hB26		1					C	C	T	G	.	.
hB27		1					C	C	T	G	.	.

Nota: se utilizó como referencia el genoma de GS (ensamblaje B). Los puntos indican posiciones conservadas respecto al genoma de referencia. Los guiones indican huecos en el alineamiento. Para cada haplotipo se indican el número de clones por aislado. Las mutaciones no sinónimas se señalan en negrita. *Haplotipos idénticos que se diferenciaron únicamente en la longitud de las secuencias.

Tabla 8.5: Haplotipos del locus *calt* ensamblaje A.

	Aislados							Posiciones variantes												
Haplotipos	122	147	152	209	251	263	321	839	72	99	118	171	289	303	318	336	356	357	358	363
WB									C	A	T	A	A	T	A	A	A	C	G	G
DH									.	.	G	.	.	.	G	A
hA1	8	8	8	1	5		5	7	.	.	G	.	.	.	G	A
hA2				3	3	7			.	.	G	.	C	.	G	A
hA3				2					.	.	G	.	C	.	G	-	-	-	-	A
hA4				1					T	.	G	.	C	.	G	A
hA5				1					.	.	C	G	.	C	.	G	.	.	.	A
hA6						1			.	.	G	G	C	.	G	A
hA7							1		.	.	G	.	.	G	G	A
hA8							1		.	G	.	G	.	.	G	A

Nota: se utilizó como referencia el genoma de WB (ensamblaje A). Los puntos indican posiciones conservadas respecto al genoma de referencia. Los guiones indican huecos en el alineamiento. Para cada haplotipo se indican el número de clones por aislado. Las mutaciones no sinónimas se señalan en negrita.

Tabla 8.6: Haplotipos del locus *calt* ensamblaje B.

Haplotipos	Aislados					Posiciones variantes																			
	122	321	407	704	1221	1343	68	69	87	162	164	189	217	219	243	270	303	320	321	332	348	354	376		
GS							A	G	T	C	A	C	C	G	C	A	C	A	G	A	C	C	C	C	
hB1	4						—	—	T	.	T	.	A	
hB2	1						—	—	T	.	T	.	A	G	
hB3		3					—	—	A	.	T	.	.	
hB4		1	2		1		.	A	A	.	T	.	.	
hB5		1					—	—	A	A	.	T	.	.	
hB6		1					—	—	G	.	.	T	.	T	
hB7		1					—	—	T	.	.	
hB8			3				.	A	A	.	T	T	.	
hB9			1				.	A	T	.	T	.	T	.	A	.	T	.	.	.	
hB10			1				.	A	T	G	T	.	A	.	T	.	.	.	
hB11			1				.	A	T	.	T	.	A	
hB12				8			—	—	T	.	A	.	T	.	.	
hB13					4		.	A	.	T	.	T	A	.	T	.	.	
hB14					1		.	A	C	T	.	T	A	.	T	.	.	
hB15					2		.	A	T	.	A	
hB16						6	A	
hB17					1		G	A	
hB18					1		G	A	

Nota: se utilizó como referencia el genoma de GS (ensamblaje B). Los puntos indican posiciones conservadas respecto al genoma de referencia. Los guiones indican huecos en el alineamiento. Para cada haplotipo se indican el número de clones por aislado. Las mutaciones no sinónimas se señalan en negrita y una mutación que introdujo un codón de parada se señala en negrita y cursiva.

8.2 TABLAS DE HAPLOTIPOS DEL EXPERIMENTO II

Tabla 8.7: Haplotipos del locus *bg*.

Alelos	Aislados					Posiciones variantes						
	D	E	H	K	M	204	210	225	285	354	369	438
GS						G	T	G	T	T	T	C
<i>bg_1</i>	8	8	8	8		A	T
<i>bg_2</i>				8		C	T
<i>bg_3</i>					8	.	C	.	.	C	C	.
<i>bg_4</i>					8	A	C	.	.	C	C	.
<i>bg_5</i>					8	.	C	A	.	C	C	.
<i>bg_6</i>					8	.	C	.	C	C	C	.

Nota: se utilizó como referencia el genoma de GS (ensamblaje B). Los puntos indican posiciones conservadas respecto al genoma de referencia. Para cada alelo se indican el número de quistes por aislado en los que se identificó. Las mutaciones no sinónimas se señalan en negrita.

Tabla 8.8: Haplotipos del locus *tpi*.

	<u>Aislados</u>					<u>Posiciones variantes</u>									
Alelos	D	E	H	K	M	39	51	67	91	165	168	210	250	304	
GS						A	G	A	T	T	T	A	C	A	
<i>tpi_1</i>	8	8	8	8		
<i>tpi_2</i>				8		G	.	.	C	C	C	G	.	G	
<i>tpi_3</i>					8	G	.	.	C	C	C	G	.	.	
<i>tpi_4</i>					8	.	A	G	C	.	C	G	.	G	
<i>tpi_5</i>					1	G	.	.	C	C	C	G	T	G	

Nota: se utilizó como referencia el genoma de GS (ensamblaje B). Los puntos indican posiciones conservadas respecto al genoma de referencia. Para cada alelo se indican el número de quistes por aislado en los que se identificó. Las mutaciones no sinónimas se señalan en negrita.

Tabla 8.9: Haplotipos del locus *rpL7a*.

Alelos	Aislados					Posiciones variantes									
	D	E	H	K	M	288	297	348	363	405	420	434	456	505	563
GS						C	C	C	G	A	C	C	C	C	G
<i>rpL7a_1</i>	8		1		8
<i>rpL7a_2</i>	8					<i>T</i>	.	.	.
<i>rpL7a_3</i>		8	8		8	.	<i>T</i>
<i>rpL7a_4</i>					8	<i>G</i>
<i>rpL7a_5</i>						7	<i>T</i>	.	<i>T</i>	.
<i>rpL7a_6</i>						1	<i>T</i>	.	<i>T</i>	<i>A</i>	.	<i>T</i>	.	.	<i>T T A</i>

Nota: se utilizó como referencia el genoma de GS (ensamblaje B). Los puntos indican posiciones conservadas respecto al genoma de referencia. Para cada alelo se indican el número de quistes por aislado en los que se identificó. Las mutaciones no sinónimas se señalan en negrita. En cursiva se indican las posiciones intrónicas.

Tabla 8.10: Haplotipos del locus *mlh1*.

Alelos	Aislados					Posiciones variantes														
	D	E	H	K	M	247	254	270	291	303	306	309	315	318	359	365	375	420	426	440
GS						A	A	C	C	A	T	C	T	G	C	T	C	A	G	A
<i>mlh1_1</i>	8	1		8		<i>C</i>	.	.	A	.
<i>mlh1_2</i>	6					A	.	.
<i>mlh1_3</i>		8			8
<i>mlh1_4</i>					8	<i>G G</i>	.	<i>T G C</i>	<i>G</i>	.	.	.	<i>G C G C</i>	.	.	.
<i>mlh1_5</i>					8	<i>T</i>	.	A	.	.	.	A	.	.
<i>mlh1_6</i>					8	<i>G</i>	.	<i>T</i>	.	<i>G</i>	.	.	C	A	<i>G</i>	.	.	<i>G C G C T A C</i>	.	.
<i>mlh1_7</i>					8	<i>G</i>	.	<i>T</i>	.	<i>G</i>	.	.	C	A	<i>G</i>	.	<i>T G C G C T A</i>	.	.	.
<i>mlh1_8</i> [†]						<i>G</i>	.	<i>T</i>	.	<i>G</i>	.	.	C	A	<i>G</i>	.	<i>T G C G C T A C</i>	.	.	.

Nota: se utilizó como referencia el genoma de GS (ensamblaje B). Los puntos indican posiciones conservadas respecto al genoma de referencia. Para cada alelo se indican el número de quistes por aislado en los que se identificó. Las mutaciones no sinónimas se señalan en negrita. [†] Alelo potencialmente recombinante que se detectó a mayor frecuencia que el alelo parental *mlh1_7* en el control del aislado.

Tabla 8.11: Haplotipos del locus *calt*.

Alelos	Aislados					Posiciones variantes					
	D	E	H	K	M	69	189	283	303	321	348
GS						G	C	G	C	G	C
<i>calt_1</i>	8	8	8	7		.	.	.	A	.	.
<i>calt_2</i>				7		.	T	.	T	A	T
<i>calt_3</i>				6		A	T	T	T	A	T
<i>calt_4</i>				8		A	T	.	T	A	T
<i>calt_5</i>				7		.	.	.	A	T	.
<i>calt_6</i>				6		A	T	.	T	A	.

Nota: se utilizó como referencia el genoma de GS (ensamblaje B). Los puntos indican posiciones conservadas respecto al genoma de referencia. Para cada alelo se indican el número de quistes por aislado en los que se identificó. Las mutaciones no sinónimas se señalan en negrita.

Tabla 8.12: Haplotipos del locus *fd*.

Alelos	Aislados					Posiciones Variantes														
	D	E	H	K	M	90	138	150	153	213	225	228	267	306	342	348	393	402	404	407
GS						G	C	G	A	T	A	G	C	G	A	A	C	T	C	A
<i>fd_1</i>	8		8			T	.	.	.	C	G	A	.	A	G	.	.	C	.	G
<i>fd_2</i>		8				.	T	.	G
<i>fd_3</i>				8		T	.	.	.	C	G	A	.	A	G	.	.	C	T	G
<i>fd_4</i>				1		.	.	A	.	C	.	.	T	.	.	.	T	.	.	G
<i>fd_5</i>				8		G	G
<i>fd_6</i>				8		G	G

Nota: se utilizó como referencia el genoma de GS (ensamblaje B). Los puntos indican posiciones conservadas respecto al genoma de referencia. Para cada alelo se indican el número de quistes por aislado en los que se identificó. Las mutaciones no sinónimas se señalan en negrita.

Tabla 8.13: Haplotipos del locus *gdh*.

Alelos	Aislados					Posiciones variantes											
	D	E	H	K	M	429	447	551	561	570	597	612	690	699	702	705	756
GS						C	C	A	T	C	C	A	G	C	C	T	T
<i>gdh_1</i>	8			8		.	.	.	C	T	T	.	A	.	.	C	C
<i>gdh_2</i>		8	8	8	
<i>gdh_3</i>				8		T	.	C
<i>gdh_4</i>				8		T	T	.	C	.	.	G	.	T	.	.	.
<i>gdh_5</i>	1					.	.	G

Nota: se utilizó como referencia el genoma de GS (ensamblaje B). Los puntos indican posiciones conservadas respecto al genoma de referencia. Para cada alelo se indican el número de quistes por aislado en los que se identificó. Las mutaciones no sinónimas se señalan en negrita.

8.3 NÚMERO DE LECTURAS POR ALELO

Tabla 8.14: Lecturas por alelo en los quistes del aislado D.

Alelos	Quistes																										
	D1			D2			D3			D4			D5			D10			D11			D13			D		
	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
<i>bg_1</i>	845	1471	3747	1871	3024	8123	435	849	2102	905	1495	3995	954	1740	4357	582	1103	2629	451	895	2164	144	328	877	1154	1938	5631
<i>bg_2</i>																											
<i>bg_3</i>																											
<i>bg_4</i>																											
<i>bg_5</i>																											
<i>bg_6</i>																											
<i>tpi_1</i>	4940	7628	8419	2082	3289	4096	1067	1794	2035	1978	3176	3798	4095	6586	7951	258	458	543	691	1182	1374	6804	10813	8520	1040	1720	1981
<i>tpi_2</i>																											
<i>tpi_3</i>																											
<i>tpi_4</i>																											
<i>tpi_5</i>																											
<i>rpL7a_1</i>	174	338	599	230	425	706	107	203	305	257	450	750	645	1024	1728	175	309	493	192	175	484	1126	1771	2872	321	562	970
<i>rpL7a_2</i>	124	202	221	41	94	92	15	31	33	44	87	86	82	147	151	28	53	59	32	64	57	154	259	286	56	116	115
<i>rpL7a_3</i>																											
<i>rpL7a_4</i>																											
<i>rpL7a_5</i>																											
<i>rpL7a_6</i>																											
<i>mlh1_1</i>	19	56	52	14	55	44	8	25		19	62	69	57	176	162	14	55		9	28		428	1005	1089	72	229	210
<i>mlh1_2</i>	3	33	8	7	21	8	8	29		4	45	8	10	78	33				2	18		9	67	23	15	91	41
<i>mlh1_3</i>																											
<i>mlh1_4</i>																											
<i>mlh1_5</i>																											
<i>mlh1_6</i>																											
<i>mlh1_7</i>																											
<i>mlh1_8</i>																											
<i>gdh_1</i>	65	348	843	78	243	956	36	137	458	66	198	761	212	569	2355	38	136	420	63	195	697	76	353	816	51	165	509
<i>gdh_2</i>																		16	7								
<i>gdh_3</i>																											
<i>gdh_4</i>																											
<i>gdh_5</i>																											
<i>calt_1</i>	35	92	80	10	46		13	31		12	43		39	125	65	2	16		11	18		170	410	343	192	539	342
<i>calt_2</i>																											
<i>calt_3</i>																											
<i>calt_4</i>																											
<i>calt_5</i>																											
<i>calt_6</i>																											
<i>fd_1</i>	2		51		26						16		9	96	157		18		23		2	39	55		9	37	39
<i>fd_2</i>																											
<i>fd_3</i>																											
<i>fd_4</i>																											
<i>fd_5</i>																											
<i>fd_6</i>																											

Nota: para cada quiste se indica el número de lecturas asignadas a cada alelo por: A) USEQ, B) AMPLIAN, C) AMPLISAS. En rojo se señalan los casos en que un alelo se detectó en un quiste, pero no alcanzó el umbral de frecuencia del 7,5%. En amarillo se señalan lecturas que tuvieron que ser reasignadas manualmente a uno de los alelos tras haber sido asignadas por AMPLISAS a un haplotipo con un error de alineamiento.

Tabla 8.15: Lecturas por alelo en los quistes del aislado E.

Alelo	Quistes																										
	E1			E5			E6			E7			E8			E9			E12			E13			E		
	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
<i>bg_1</i>	2186	3666	8660	932	1767	4571	1765	3004	8171	439	832	2016	710	1255	3205	998	1736	4306	1317	2461	6432	3935	6576	8389	969	1563	3671
<i>bg_2</i>																											
<i>bg_3</i>																											
<i>bg_4</i>																											
<i>bg_5</i>																											
<i>bg_6</i>																											
<i>tpi_1</i>	3001	5065	5797	1781	2886	3472	1979	3441	4093	984	1676	1966	2001	3406	3984	2251	3687	4427	3762	6245	7377	4136	6846	7888	1199	1829	2151
<i>tpi_2</i>																											
<i>tpi_3</i>																											
<i>tpi_4</i>																											
<i>tpi_5</i>																											
<i>rpL7a_1</i>																											
<i>rpL7a_2</i>																											
<i>rpL7a_3</i>	997	1481	2416	656	969	1481	732	1143	1652	651	1014	1573	490	707	1098	586	982	1503	687	1089	1818	793	1166	1837	209	367	563
<i>rpL7a_4</i>																											
<i>rpL7a_5</i>																											
<i>rpL7a_6</i>																											
<i>mlh1_1</i>	26	113	63																						2	23	7
<i>mlh1_2</i>																											
<i>mlh1_3</i>	101	339	243	111	298	234	99	306	234	80	241	183	118	273	235	174	448	373	340	980	753	312	819	671	39	131	95
<i>mlh1_4</i>																											
<i>mlh1_5</i>																											
<i>mlh1_6</i>																											
<i>mlh1_7</i>																											
<i>mlh1_8</i>																											
<i>gdh_1</i>																											
<i>gdh_2</i>	571	1066	2403	337	629	1326	250	539	1260	246	468	1128	252	485	1143	317	695	1332	306	599	1498	237	478	1005	155	305	639
<i>gdh_3</i>																											
<i>gdh_4</i>																											
<i>gdh_5</i>													19	45	63												
<i>calt_1</i>	48	124	78	49	152	81	56	141	94	44	118	70	27	105	62	68	187	135	41	145	85	29	101	56	102	242	208
<i>calt_2</i>																											
<i>calt_3</i>																											
<i>calt_4</i>																											
<i>calt_5</i>																											
<i>calt_6</i>																											
<i>fd_1</i>																											
<i>fd_2</i>	6	107		3	85		10	125	48	6	87		3	53		9	79		4	68		2	44		30		
<i>fd_3</i>																											
<i>fd_4</i>																											
<i>fd_5</i>																											
<i>fd_6</i>																											

Nota: ver Tabla 8.14

Tabla 8.16: Lecturas por alelo en los quistes del aislado H.

Alelos	Quistes																										
	H2			H3			H4			H6			H8			H9			H10			H11			H		
A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	
<i>bg_1</i>	1551	2745	7660	1575	2687	7613	1502	2557	7261	1758	2826	7105	1045	1780	4384	1175	1918	2088	658	1130	2910	636	1122	3167	910	1540	4393
<i>bg_2</i>																											
<i>bg_3</i>																											
<i>bg_4</i>																											
<i>bg_5</i>																											
<i>bg_6</i>																											
<i>tpi_1</i>	5562	9032	8564	5700	9029	8579	1194	1997	2315	5449	8636	8510	2240	3640	4196	2593	4174	5050	1480	2379	2896	2555	4153	4907	681	1111	1265
<i>tpi_2</i>																											
<i>tpi_3</i>																											
<i>tpi_4</i>																											
<i>tpi_5</i>																											
<i>rpL7a_1</i>				76	133	332																					
<i>rpL7a_2</i>																											
<i>rpL7a_3</i>	899	1423	2362	913	1345	2228	568	868	1323	1029	1533	2408	518	747	1089	487	784	1195	510	780	1240	441	706	1152	303	466	681
<i>rpL7a_4</i>																											
<i>rpL7a_5</i>																											
<i>rpL7a_6</i>																											
<i>mlh1_1</i>	172	478	419	164	438	400	87	256	211	119	331	317	28	84	88	57	142	128	57	140	126	42	122	103	42	109	105
<i>mlh1_2</i>																											
<i>mlh1_3</i>																											
<i>mlh1_4</i>																											
<i>mlh1_5</i>																											
<i>mlh1_6</i>																											
<i>mlh1_7</i>																											
<i>mlh1_8</i>																											
<i>gdh_1</i>																											
<i>gdh_2</i>	608	1138	2693	438	1039	2141	329	674	1526	574	1069	2513	239	740	990	257	498	1123	317	667	1386	230	411	929	177	331	778
<i>gdh_3</i>																											
<i>gdh_4</i>																											
<i>gdh_5</i>																											
<i>calt_1</i>	54	168	107	48	134	98	18	58	50	51	133	95	18	65		17	51		12	51		17	74		54	126	97
<i>calt_2</i>																											
<i>calt_3</i>																											
<i>calt_4</i>																											
<i>calt_5</i>																											
<i>calt_6</i>																											
<i>fd_1</i>	10	67	104	16	101	122	9	47	63	2	35	96	2	23		2	23		19	40		19		6	33		
<i>fd_2</i>																											
<i>fd_3</i>																											
<i>fd_4</i>																											
<i>fd_5</i>																											
<i>fd_6</i>																											

Nota: ver Tabla 8.14

Tabla 8.17: Lecturas por alelo en los quistes del aislado K.

Alelo	Quistes																										
	K3			K6			K8			K21			K25			K26			K27			K28			K		
	A	B	*C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
<i>bg_1</i>	274	558	1557	683	1240	3441	470	891	2376	485	899	2588	541	921	2581	513	950	2970	201	372	1077	368	870	2250	259	674	2204
<i>bg_2</i>	425	774	1217	457	730	1203	1153	1809	3025	655	1180	2187	683	1123	1716	494	914	1362	246	488	706	1463	2421	4470	442	931	1975
<i>bg_3</i>																											
<i>bg_4</i>																											
<i>bg_5</i>																											
<i>bg_6</i>																											
<i>quimera</i>	169			255				480	479		304	462	143	351	457	134	389	560	63	221	203		613	726	161	550	947
<i>quimera</i>								274					153	306	408	122	255		53	113		342		139	302	587	
<i>tpi_1</i>	195	458	515	551	1174	1771	408	884	1272	742	1523	2173	490	1027	1375	310	734	1050	236	599	846	403	972	1490	111	297	382
<i>tpi_2</i>	265	539	627	797	1534	2121	891	1673	2256	305	630	904	211	427	591	426	848	1230	191	496	589	1297	2377	3486	137	265	423
<i>tpi_3</i>																											
<i>tpi_4</i>																											
<i>tpi_5</i>																											
<i>quimera</i>																				117							
<i>quimera</i>								289			247		182			220	226		193	128		434	485		95	95	
<i>rpL7a_1</i>																											
<i>rpL7a_2</i>																											
<i>rpL7a_3</i>	156	277	344	242	445	638	372	612	812	1206	1951	2743	337	544	836	266	447	781	23	51	76	242	421	615	121	217	340
<i>rpL7a_4</i>	146	222	350	327	520	841	222	346	475	29	78	95	52	91	146	143	289	450	140	204	347	183	314	558	93	158	271
<i>rpL7a_5</i>																											
<i>rpL7a_6</i>																											
<i>quimera</i>																			20	58		80		31		67	
<i>quimera</i>																								32		92	
<i>mlh1_1</i>																											
<i>mlh1_2</i>																											
<i>mlh1_3</i>	34	187	87	19	68	35	22	90	54	81	385	187	52	183	128	33	211	77	33	130	59	58	329	127	21	105	40
<i>mlh1_4</i>	92	191	161	38	121	69	96	195	163	155	431	269	37	127	70	156	419	294	44	139	85	143	321	257	46	111	80
<i>mlh1_5</i>																											
<i>mlh1_6</i>																											
<i>mlh1_7</i>																											
<i>mlh1_8</i>																											
<i>gdh_1</i>	4	27	64	15	64	118	18	82	163	13	70	117	17	87	183	7	65	74	3	46	64	22	111	210	7	48	73
<i>gdh_2</i>	31	129	163	56	186	264	123	327	512	124	351	693	57	173	263	150	392	654	31	127	189	82	265	318	48	155	222
<i>gdh_3</i>																											
<i>gdh_4</i>																											
<i>gdh_5</i>																											
<i>quimera</i>																	50		20								
<i>calt_1</i>				4			3	10		9	60	16	8	40	16	6	53	14	3	23		12		4	28		
<i>calt_2</i>	4	14					15	21		16	86	29	19	77	38	23	87	39	4	12				5	30		
<i>calt_3</i>		5									88			17			14					61		19			
<i>calt_4</i>																											
<i>calt_5</i>																											
<i>calt_6</i>																											
<i>quimera</i>													14														
<i>quimera</i>													2	19	5	2		5	2								
<i>quimera</i>															2												
<i>quimera</i>																									2		
<i>fd_1</i>																											
<i>fd_2</i>																											
<i>fd_3</i>		21		16			2	18		3	15		2	11		2	18		3	21		3	30	71		10	
<i>fd_4</i>		5																									
<i>fd_5</i>																											
<i>fd_6</i>																											

Nota: Ver Tabla 8.14

Tabla 8.18: Lecturas por alelo en los quistes del aislado M.

Alelo	Quistes																											
	M14			M15			M16			M17			M19			M22			M23			M24			M			
A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C		
bg_1																												
bg_2																												
bg_3	1159	2155	2433	1002	1901	2693	916	1806	1766	317	581	784	586	1152	1052	550	1042	1319	484	1087	1310	197	439	364	354	755	1064	
bg_4	1221	2007	2521	740	1308	1195	437	794	743	155	275	251	156	313	307	350	631	537	464	984	775	255	519	562	308	641	545	
bg_5	471	878	926	411	765	680	565	1009	1051	212	402	365	653	1144	1208	94	212	165	653	1286	1093	332	699	601	182	388	363	
bg_6	310	568	626	595	1036	953	1689	2673	4167	188	317	306	1344	2121	3200	221	441	357	382	897	672	95	228	179	174	353	331	
tpi_1																												
tpi_2																												
tpi_3	1970	3327	4752	2415	4444	5551	3140	5053	5844	564	1094	1357	2362	3994	5455	284	546	740	309	594	714	503	969	1229	135	271	417	
tpi_4	331	871	1061	308	827	989	449	981	1158	157	376	439	232	638	734	115	276	328	151	386	371	81	234	266	69	199	263	
tpi_5																			32	106	105							
quimera																											67	
quimera																	116						127					
quimera																											69	
rpL7a_1	170	338	578	346	704	1130	260	550	993	37	96	155	325	616	988	62	99	179	211	379	534	70	173	247	79	155	271	
rpL7a_2																												
rpL7a_3																												
rpL7a_4																												
rpL7a_5	98	161	236	153	299	421	138	227	415	40	89	126	77	148	233	41	86	106		5		29	73	77	27	58	78	
rpL7a_6	2	25		4			10	51	33				8	47	19				4	37	24					6		
quimera																									18		31	
mlh1_1																												
mlh1_2																												
mlh1_3																												
mlh1_4																												
mlh1_5	19	84	43	28	103	56	49	210	105	4	21		26	161	70	3	22		13	80	30	4	41	16	5	45	10	
mlh1_6	64	145	100	127	313	219	72	155	105	4	19		66	171	121	2	16		32	87	60	12	31	22	17	53	27	
mlh1_7	11	38	23	20	55	39	39	124	79	3	8		56	153	123	2	8		4	16	11	10	31	18				
mlh1_8	4			5			4						8						2		3				2		4	
quimera														79									11					
quimera																						2						
quimera																						2						
gdh_1																												
gdh_2																												
gdh_3	89	284	281	146	456	429	75	237	223	32	151	82	125	390	318	71	197	176	32	163	101	38	205	130	15	78	57	
gdh_4	4	57	53	10	128	180	3	48	58	4	50	57	8	131	117				10	70	87	7	61	55		42		
gdh_5																												
quimera																			3	30								
quimera																										2		
quimera																	18			29						3		
quimera											24																5	
calt_1																												
calt_2																												
calt_3																												
calt_4		4			13			22						42													34	
calt_5	2	6		4	12		8	29		2	7			29													8	
calt_6					6			7						16						5							7	
fd_1																												
fd_2																												
fd_3																												
fd_4																												
fd_5		8		10	54	26		34			12		2	42	26		8			6			13				16	
fd_6		11		5	76	25	3	38			10			47	25		8			13			16				8	
quimera					12									10														

Nota: Ver Tabla 8.14

8.4 FIGURAS DE LAS FRECUENCIAS ALÉLICAS RELATIVAS EN LOS QUISTES INDIVIDUALES DE LOS AISLADOS D, E, H, K, M

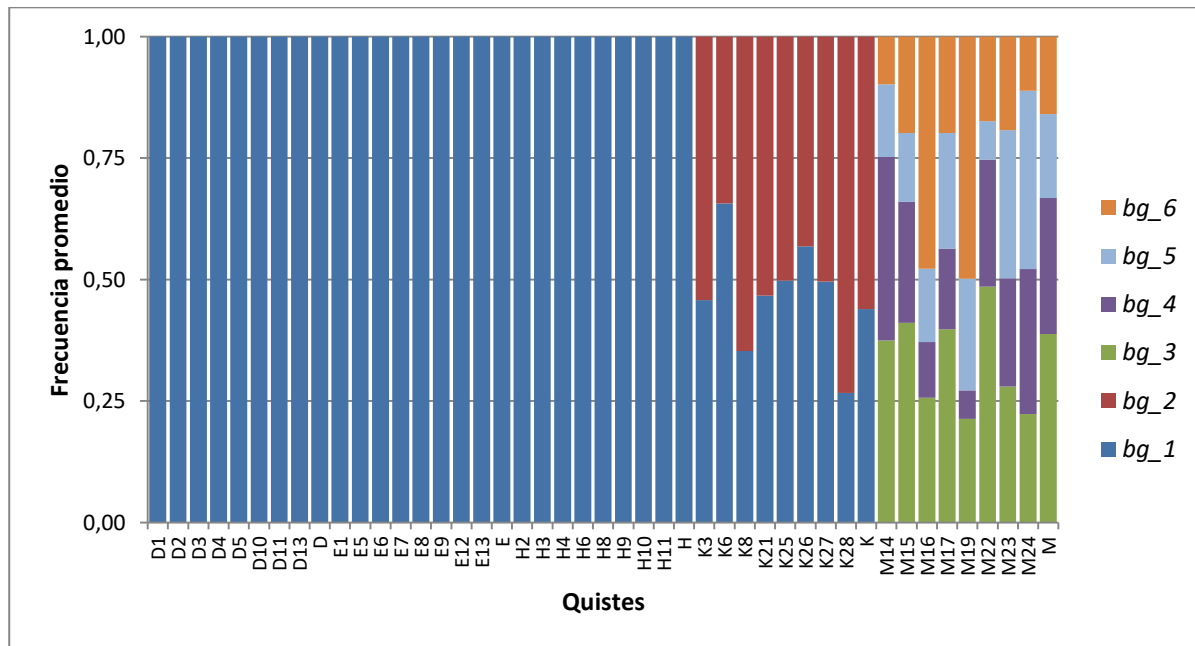


Figura 8.1: Frecuencias alélicas del locus *bg* en quistes individuales. Se muestra el promedio de las estimas de las frecuencias relativas de cada alelo obtenidas a partir del número de lecturas en USEQ, AMPLIAN y AMPLISAS. Los alelos se nombraron como en la Tabla 8.7

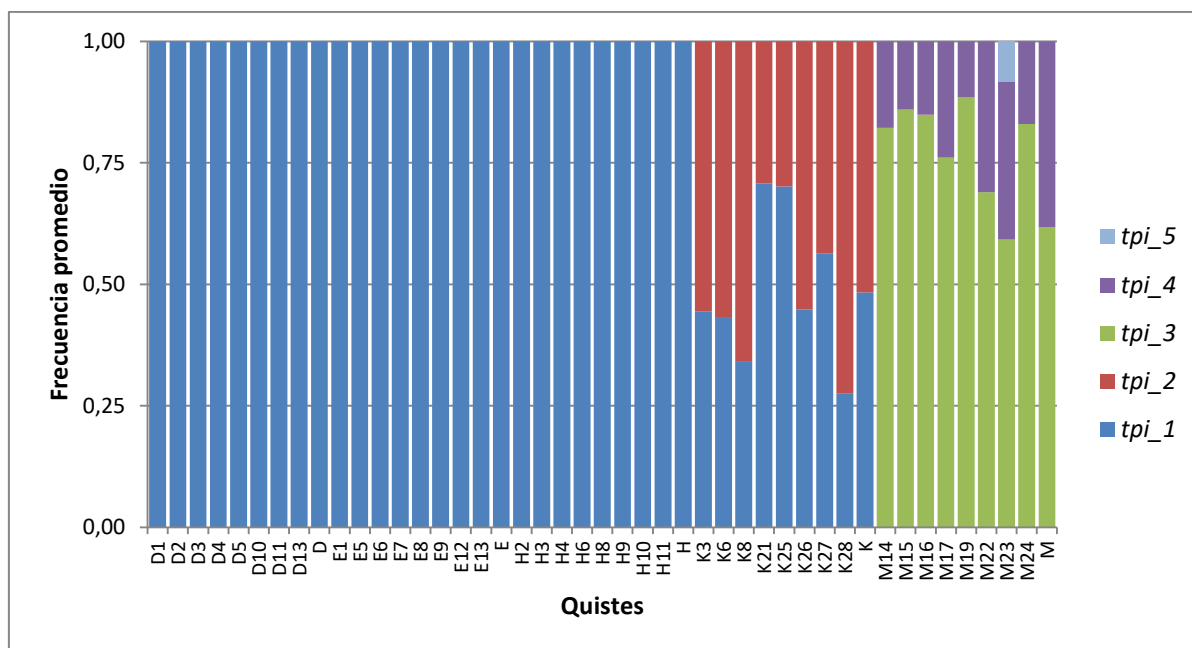


Figura 8.2: Frecuencias alélicas del locus *tpi* en quistes individuales. Se muestra el promedio de las estimas de las frecuencias relativas de cada alelo obtenidas a partir del número de lecturas en USEQ, AMPLIAN y AMPLISAS. Los alelos se nombraron como en la Tabla 8.8.

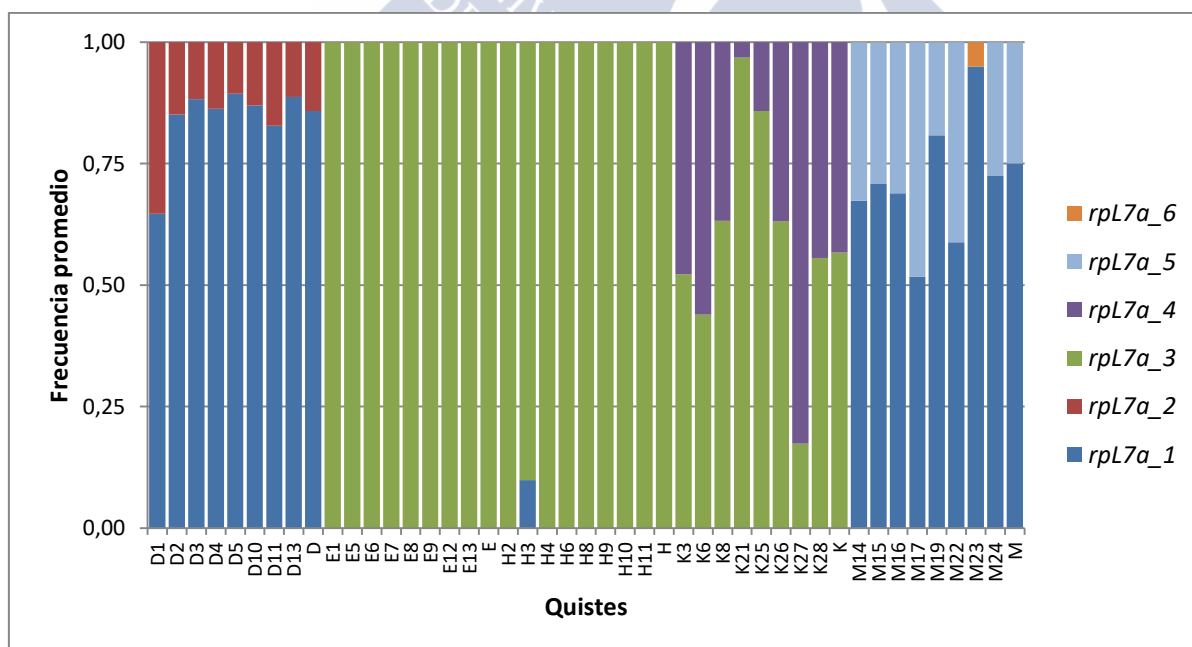


Figura 8.3: Frecuencias alélicas del locus *rpl7a* en quistes individuales. Se muestra el promedio de las estimas de las frecuencias relativas de cada alelo obtenidas a partir del número de lecturas en USEQ, AMPLIAN y AMPLISAS. Los alelos se nombraron como en la Tabla 8.9.

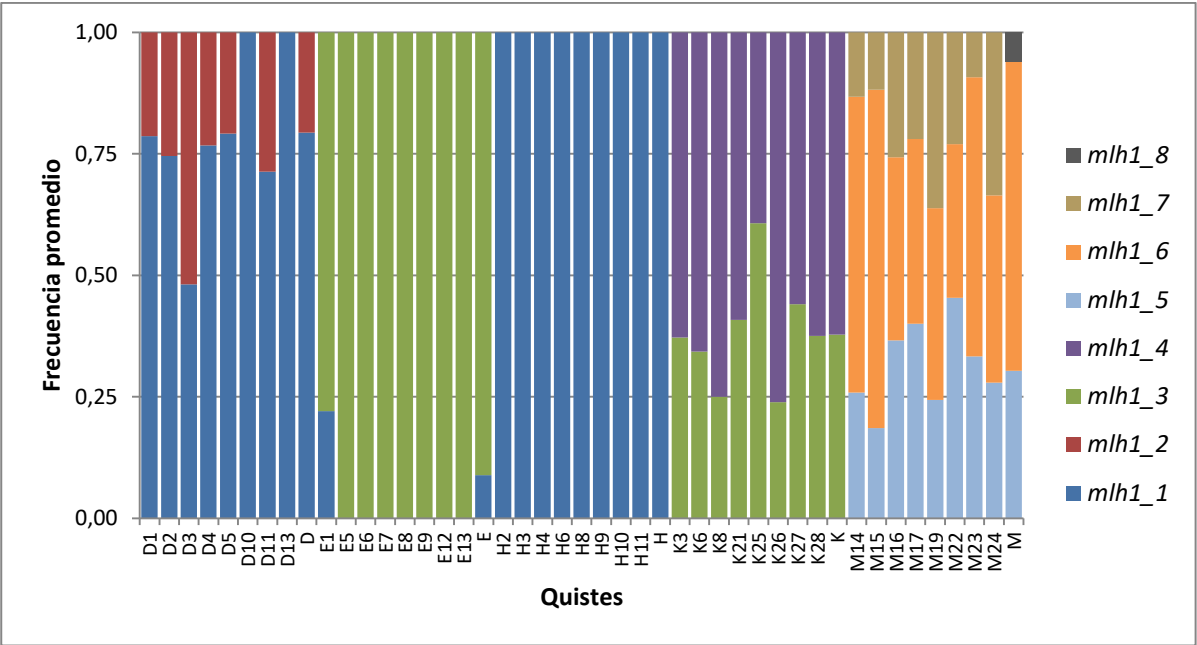


Figura 8.4: Frecuencias alélicas del locus *mlh1* en quistes individuales. Se muestra el promedio de las estimas de las frecuencias relativas de cada alelo obtenidas a partir del número de lecturas en USEQ, AMPLIAN y AMPLISAS. Los alelos se nombraron como en la Tabla 8.10.

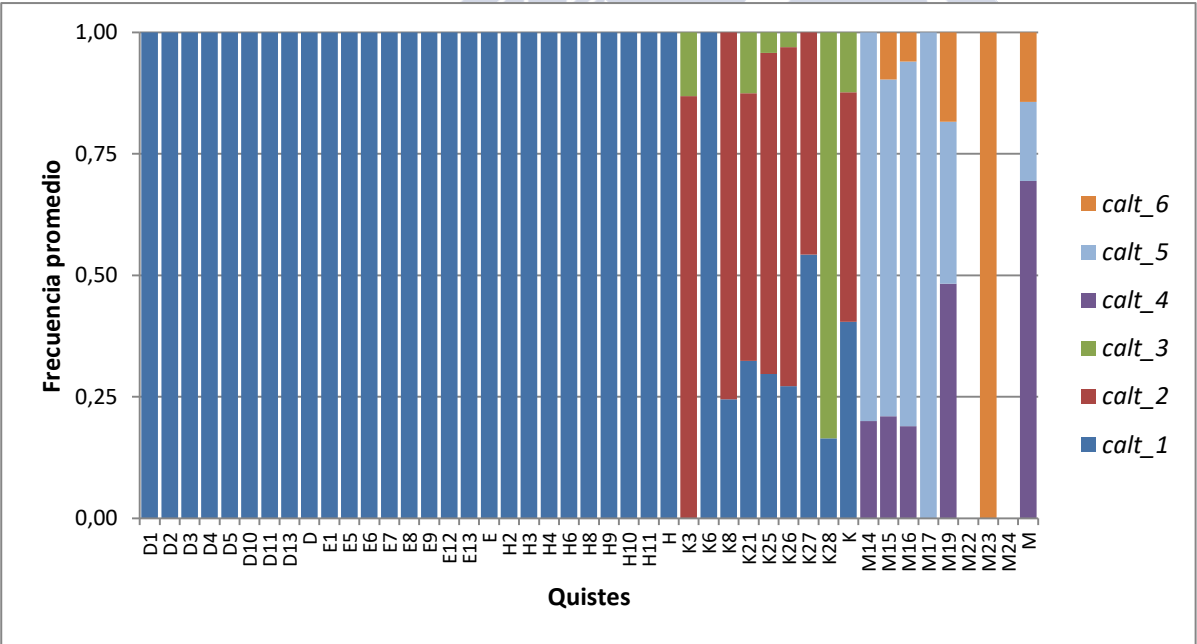


Figura 8.5: Frecuencias alélicas del locus *calt* en quistes individuales. Se muestra el promedio de las estimas de las frecuencias relativas de cada alelo obtenidas a partir del número de lecturas en USEQ, AMPLIAN y AMPLISAS. Los alelos se nombraron como en la Tabla 8.11.

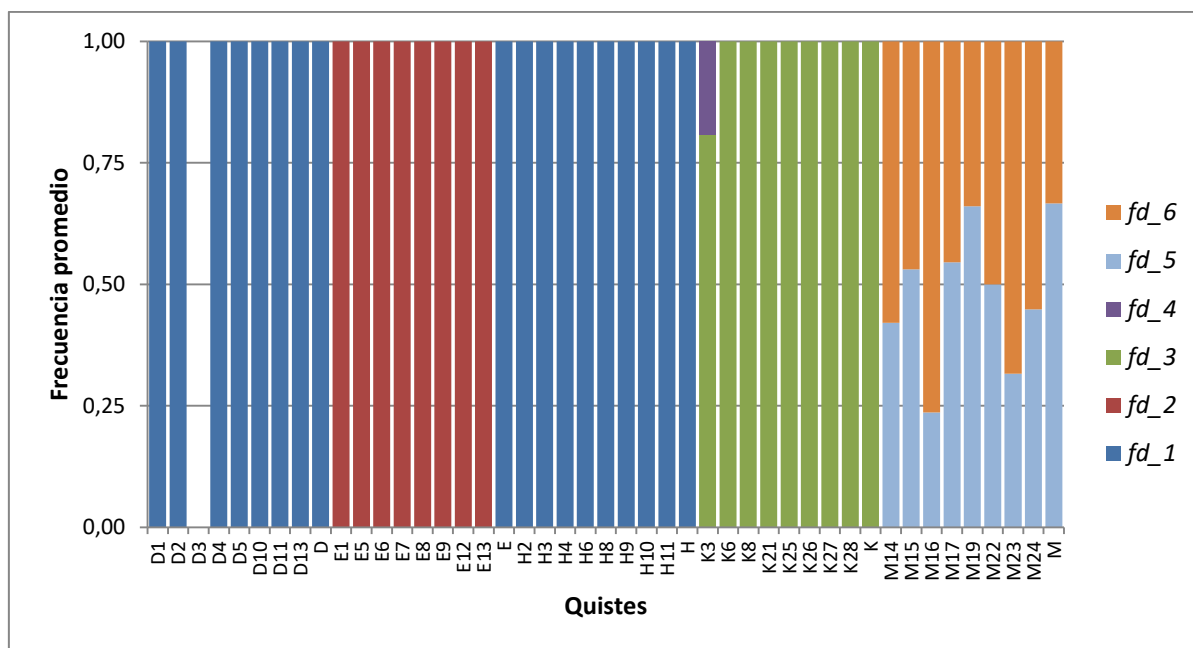


Figura 8.6: Frecuencias alélicas del locus *fd* en quistes individuales. En el eje X se indican los quistes individuales y sus controles. En el eje Y se representa el promedio de las estimas de las frecuencias alélicas obtenidas por USEQ, AMPLIAN y AMPLISAS. Los alelos se nombraron como en la Tabla 8.12.

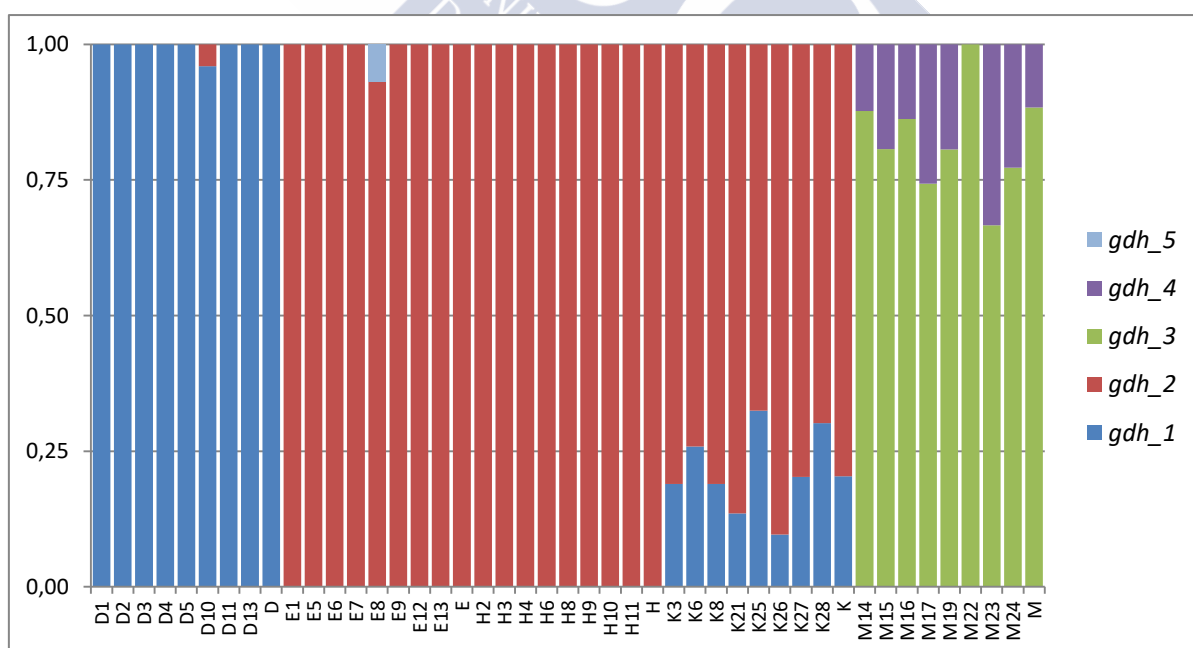


Figura 8.7: Frecuencias alélicas del locus *gdh* en quistes individuales. Se muestra el promedio de las estimas de las frecuencias relativas de cada alelo obtenidas a partir del número de lecturas en USEQ, AMPLIAN y AMPLISAS. Los alelos se nombraron como aparecen identificados en la Tabla 8.13.

8.5 ÁRBOLES FILOGENÉTICOS DEL EXPERIMENTO II

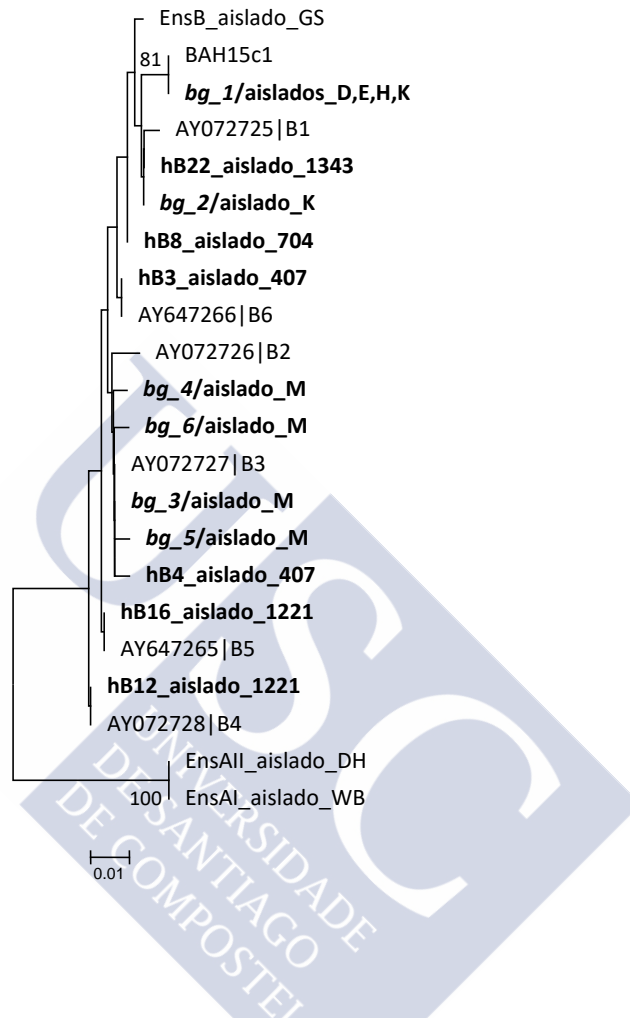


Figura 8.8: Árbol filogenético de los haplotipos del locus *bg*. Las relaciones filogenéticas fueron inferidas por el método NJ, y las distancias genéticas calculadas por el método de Tamura-Nei. Se incluyeron como referencia las secuencias de los aislados GS y BAH15c1 y las identificadas por Lalle *et al.* (2005) como genotipos B1-B6. Se utilizaron las secuencias de los genomas WB y DH del ensamblaje A como grupos externos. Los valores de *bootstrap* superiores al 75% se indican al lado de los nodos correspondientes. En negrita se indican los alelos identificados en este trabajo, así como el aislado en el que se detectaron. En los nombres de los taxones correspondientes a las secuencias de referencia se indica el número de acceso de GenBank.

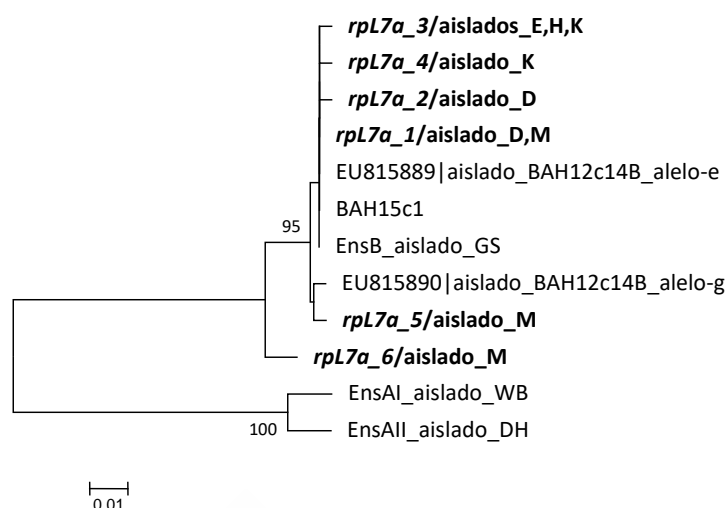


Figura 8.9: Árbol filogenético del locus *rpL7a*. Las relaciones filogenéticas fueron inferidas por el método NJ, y las distancias genéticas calculadas por el método de Tamura-Nei. Se utilizaron las secuencias de referencia de los aislados WB y DH del ensamblaje A como grupos externos. Se incluyeron como referencia las secuencias de los aislados GS BAH15c1 y BAH12c14B 1. Los valores de *bootstrap* superiores al 75% se indican al lado de los nodos correspondientes. En negrita se indican los alelos identificados en este trabajo, así como el aislado en el que se detectaron. En los nombres de los taxones correspondientes a las secuencias de referencia se indica el número de acceso de GenBank.

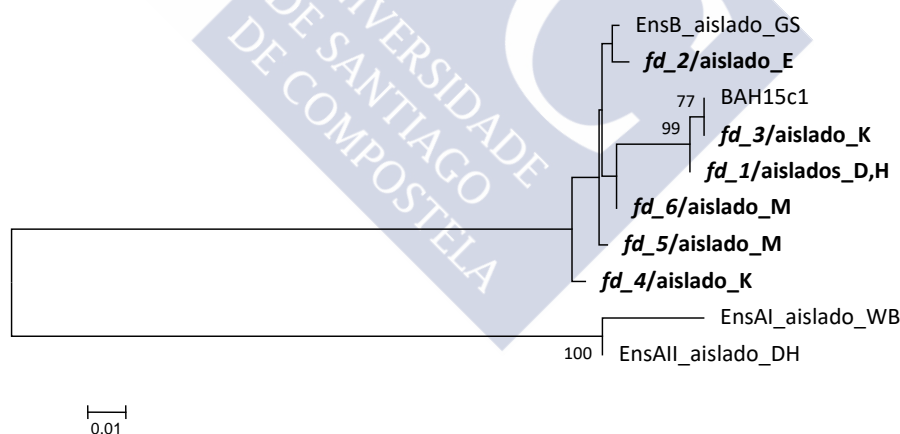


Figura 8.10: Árbol filogenético del locus *fd*. Las relaciones filogenéticas fueron inferidas por el método NJ, y las distancias genéticas calculadas por el método de Tamura-Nei. Se utilizaron las secuencias de referencia de los aislados WB y DH del ensamblaje A como grupos externos. Se incluyeron como referencia las secuencias de los aislados GS, BAH15c1 y BAH12c14B. Los valores de *bootstrap* superiores al 75% se indican al lado de los nodos correspondientes. En negrita se indican los alelos identificados en este trabajo, así como el aislado en el que se detectaron. En los nombres de los taxones correspondientes a las secuencias de referencia se indica el número de acceso de GenBank.



Figura 8.11: Árbol filogenético del locus *calt*. Las relaciones filogenéticas fueron inferidas por el método NJ, y las distancias genéticas calculadas por el método de Tamura-Nei. Se utilizaron las secuencias de referencia de los aislados WB y DH del ensamblaje A como grupos externos. Se incluyeron como referencia las secuencias de los aislados GS, BAH15c1 y BAH12c14B. Los valores de *bootstrap* superiores al 75% se indican al lado de los nodos correspondientes. En negrita se indican los alelos identificados en este trabajo, así como el aislado en el que se detectaron. En los nombres de los taxones correspondientes a las secuencias de referencia se indica el número de acceso de GenBank.

